



UNIVERSIDAD DE NAVARRA  
FACULTAD ECLESIASTICA DE FILOSOFIA

PABLO JOSÉ CONCEPCIÓN FRAGOSO

**REDUCCIONISMOS CONTEMPORANEOS  
ACERCA DE LA MENTE**

*La inmaterialidad de lo mental*

Extracto de la Tesis Doctoral presentada en la  
Facultad Eclesiástica de Filosofía de la Universidad de Navarra

PAMPLONA

1996



Ad normam Statutorum Facultatis Philosophiae Universitatis Navarrensis  
perlegimus et adprobavimus

Pampilonae, die 16 mensis octobris anni 1995

Dr. Iacobus NUBIOLA

Dr. Marianus ARTIGAS

Coram tribunali, die 22 mensis iunii anni 1994, hanc  
dissertationem ad Lauream Candidatus palam defendit

Secretarius Facultatis  
Dr. Ioseph Emmanuel ZUMAQUERO

Excerpta e Dissertationibus in Philosophia  
Vol. VI, n. 2





## PROLOGO

Al estudiar las diversas doctrinas filosóficas que buscan dar respuesta a la pregunta acerca de lo que es la mente, se descubre que todavía no parece existir una respuesta que resuelva todos los problemas que surgen al intentar conocer cuál es su realidad. Es por tanto la mente un objeto de nuestro conocimiento que se nos presenta como misterioso, o al menos como una realidad que se nos vuelve esquivia y difícil de conceptualizar. La existencia de todas estas teorías filosóficas que tienen como objeto explicar la realidad de la mente, muestra lo importante y decisiva que es la respuesta que se dé a esta pregunta, ya que en esa respuesta, está en juego la respuesta a la pregunta más existencial que puede hacerse una persona: ¿Quién o qué soy yo? De la respuesta que se dé al problema de la mente, dependerá el que se tenga, o no, una cabal comprensión de lo que es el ser humano. Esta consideración está presente a lo largo de todo el estudio, y se manifiesta en que en la tesis no he buscado sólo hacer una crítica, la crítica ha sido necesaria para mostrar que las teorías reduccionistas siempre dejan algo fuera, en cambio, el objetivo de este estudio ha sido intentar recuperar una visión unitaria del hombre que muestre que esa unidad no puede ser reconstituida analíticamente.

La metodología que he empleado ha estado marcada por este objetivo, ya que lo que realmente me ha interesado de estas teorías ha sido su contenido filosófico y cuáles eran los problemas a los que estas teorías no conseguían dar respuesta. Por eso el estudio ha sido un estudio sistemático, en el que no he pretendido realizar un análisis histórico de las teorías propuestas por los diversos filósofos de lo mental aunque a veces me haya apoyado en consideraciones de tipo histórico para intentar clarificar las diversas doctrinas filosóficas acerca de la mente. Pero si hubiera reducido el estudio a este único aspecto me habría quedado en una visión negativa del problema de la mente, en un reconocer que no existe una teoría que dé solución a los problemas que la mente plantea. Era necesario que intentara aportar luz en el ámbito del problema de la mente, para ello el primer paso que he dado fue estudiar a H. Putnam y J. R. Searle por ser dos filósofos que muestran que las teorías reduccionistas no dan respuesta al problema de la mente, al

tiempo que hacen ver la importancia que en este estudio tiene uno de los temas centrales de la teoría del conocimiento, que es el de la intencionalidad. Lo que he aspirado a hacer en esta tesis doctoral es estudiar las tesis de H. Putnam y J. R. Searle, intentando enriquecerlas con los trabajos de otros filósofos de lo mental.

En ese estudio he buscado responder a esta pregunta acerca de la mente, no tanto identificando o mostrando dos realidades en el ser humano, una de carácter material y otra espiritual, sino buscando la comprensión de su unidad; el ser humano no es sólo materia, sino que descubrimos en él una unidad más profunda que la de los seres inanimados. Hay una unidad en el hombre que aunque impide decir que la mente no posee un substrato material, al mismo tiempo impide afirmar que la mente es sólo materia informada, ya que se descubre en ella realidades que la caracterizan como inmaterial, pero no por ello debe entenderse inmaterial como si significara sin relación con la materia. Esa realidad inmaterial será consecuencia del tipo de vida que caracteriza al ser humano y que le hace poseer una dinamicidad propia que no se encuentra en los seres inanimados. Esta dinamicidad también se descubre en los demás seres animados, pero en el caso del hombre, es una dinamicidad superior por alcanzar a través de ella unas capacidades que no tienen los otros seres animados. Esto permite descubrir que el progreso en la comprensión de la realidad de la mente está muy ligado al estudio de otra pregunta, que hoy en día, a menudo, es dejada de lado, y que es la pregunta acerca de qué es la vida. Una clarificación de lo mental comporta, por tanto, una clarificación de lo que significa la vida.

En este *excerptum* he recogido el tercer y cuarto capítulo de la tesis, en los que se estudian las razones que aducen Putnam y Searle para mostrar que no se puede reducir la mente a fenómenos observables. Pero no quiere esto significar que este *excerptum* no posea una unidad propia, pero sí que es cierto que sin las otras partes de la tesis este estudio se convierte en una crítica a los reduccionismos contemporáneos, lo que conlleva un cierto carácter de negatividad. Es por eso por lo que a la hora de leerlo conviene no llegar a la conclusión de que es imposible una respuesta al problema mente-cuerpo: esa solución es posible pero para ello hay que mostrar que es necesario recuperar la *psyche* aristotélica. A esa recuperación está dedicado el último capítulo de la tesis.

No quiero terminar sin agradecer las sugerencias que D. Leonardo Polo y D. Miguel Bastons me hicieron en el examen de la tesis de licenciatura y que me han servido para enriquecer con una mayor profundidad filosófica esta tesis doctoral, a D. Jorge Vicente por las ideas que me sugirió en su curso de doctorado, y a D. Jaime Nubiola por guiarme en este difícil

camino de entender mejor lo que es la mente y no haber perdido la paciencia en ese largo recorrido que hemos hecho juntos. Por último, quiero extender mi agradecimiento y gratitud a los que de un modo u otro han colaborado en la realización de este trabajo de investigación.





## INDICE DE LA TESIS

	<u>PÁG.</u>
TABLA DE ABREVIATURAS .....	V
INTRODUCCION .....	1
1. LA FILOSOFIA DE LA MENTE .....	11
1.1. El problema mente-cuerpo .....	12
1.2. Filosofías de la mente .....	19
1.3. Interaccionismo dualista cartesiano .....	27
1.4. Interaccionismo dualista de Popper y Eccles .....	31
1.5. Interaccionismo aristotélico .....	36
1.6. Interaccionismo monista .....	38
1.7. Resumen .....	40
2. TEORIAS REDUCCIONISTAS DE LA MENTE .....	41
2.1. El término “reduccionismo” aplicado a las filosofías de la mente .....	42
2.2. Los reduccionismos de la mente .....	47
2.3. Conductismo .....	51
2.4. Teorías de la identidad de tipos .....	53
2.5. Teorías de la identidad de eventos .....	58
2.5.1. Monismo anómalo .....	60
2.5.2. Funcionalismo .....	66
2.5.2.1. El lenguaje del pensamiento .....	67
2.5.2.2. La psicología popular .....	77
2.5.3. IA fuerte o funcionalismo computacional .....	80
2.6. Epifenomenalismo .....	84
2.7. Eliminativismo .....	87
2.7.1. Materialismo eliminativo .....	87

2.7.2. El instrumentalismo .....	95
2.7.3. El instrumentalismo y la conciencia .....	100
2.8. Resumen .....	104
<b>3. LAS CRITICAS DE HILARY PUTNAM .....</b>	<b>107</b>
3.1. Introducción .....	107
3.2. La intencionalidad es primitiva .....	109
3.3. La intencionalidad no es reductible .....	132
3.3.1. El significado no es reductible .....	133
3.3.2. El funcionalismo no funciona .....	138
3.4. La intencionalidad es ineliminable .....	150
3.5. Resumen .....	160
<b>4. CRITICAS DE JOHN R. SEARLE .....</b>	<b>165</b>
4.1. Introducción .....	165
4.2. Los reduccionismos de la mente se basan en una ontología de la tercera persona .....	166
4.3. La mente no es un ordenador digital .....	192
4.4. Resumen .....	207
<b>5. EL ALMA ARISTOTELICA .....</b>	<b>213</b>
5.1. Introducción .....	213
5.2. Interpretaciones de Aristóteles .....	215
5.2.1. Aristóteles y el cartesianismo .....	217
5.2.2. Aristóteles y el materialismo .....	228
5.2.3. Aristóteles y el funcionalismo .....	232
5.3. La visión aristotélica del alma .....	238
5.3.1. Los términos que emplea Aristóteles .....	238
5.3.2. La psicología de Aristóteles sobre la percepción .....	246
5.4. La teoría aristotélica en nuestros días .....	249
5.4.1. La mente y el cuerpo .....	249
5.5. Resumen .....	265
<b>CONCLUSIONES .....</b>	<b>267</b>
<b>BIBLIOGRAFIA .....</b>	<b>281</b>





## BIBLIOGRAFIA DE LA TESIS

Se relacionan a continuación, ordenados alfabéticamente, los libros y artículos consultados y citados a lo largo de la Tesis.

- ACKRILL, J. L., *La filosofía de Aristóteles*, Caracas, Monte Avila Editores, trad. esp. de F. Bravo, 1984.
- AGAZZI, E. (ed.), *The Problem of Reductionism in Science*, Dordrecht, Kluwer Academic Publishers, 1991
- ALLEN, C., "Mental Content and Evolutionary Explanation", en *Biology and Philosophy*, VII, 1992, pp. 1-12.
- ANNAS, J., "Aristotle on Memory and the Self", en M. C. Nussbaum y A. O. Rorty (eds.) *Essays on Aristotle's De Anima*, Oxford, Clarendon Press, 1992.
- ANSCOMBE, G. E. M., "The Causation of Action", en C. Ginet y S. Shoemaker (eds.) *Knowledge and Mind. Philosophical Essays*, Oxford, Oxford University Press, 1983.
- ARENS, H., *Aristotle's Theory of Language and its Tradition*, Amsterdam, John Benjamins, 1984.
- ARISTOTELES, *De la Generation des Animaux*, por P. Louis (ed. lit.), París, Belles Lettres, 1951.
- ARISTOTELES, *De l'Ame*, por A. Jannone (ed. lit.) y E. Barbotin (trad.), París, Belles Lettres, 1966.
- ARISTOTELES, *De interpretatione*, Valencia, Teorema, Trad. esp. de A. García Suárez y J. Velarde Lombraña, 1977.
- ARISTOTELES, *Acerca del alma*, Madrid, Gredos, trad. esp. de T. Calvo Martínez, 1978.
- ARISTOTELES, *Metafísica*, Madrid, Gredos, trad. esp. de V. García Yebra, 2ª Ed., 1982.
- ARISTOTELES, *Etica a Nicómaco*, Madrid, Gredos, trad. esp. de J. Pallí Bonet, 1988.

- ARISTOTELES, *Parva Naturalia*, en *The Works of Aristotle*, Vol. III, Oxford, Clarendon Press, trad. ingl. de J. I. Beare y G. R. T. Ross, 1963.
- ARISTOTELES, *The Physics*, Vol. I, London, William Heinemann, trad. ingl. de Ph. H. Wicksteed y F. M. Cornford, 1963.
- ARISTOTELES, *De Generatione Animalia*, en *The Works of Aristotle*, Oxford, Clarendon Press, trad. ingl. de A. Platt, 1965.
- ARMSTRONG, D. M., *A Materialist Theory of Mind*, London, Routledge and Kegan Paul, 1968.
- ARMSTRONG, D. M., *The Nature of Mind*, Queensland, University of Queensland Press, 1980.
- ARTIGAS, M., *La inteligibilidad de la naturaleza*, Pamplona, Eunsa, 1992.
- BAKER, L. R., "Eliminativism and an Argument from Science", en *Mind & Language*, VIII, 1993, pp. 180-188.
- BARNES, J., "Aristotle's Concept of Mind", en *Proceedings of the Aristotelian Society*, LXXII, 1971-2, pp. 101-114.
- BARREAU, H., *Aristóteles y el análisis del saber*, Madrid, EDAF, trad. esp. de I. Gómez Romero, 1977.
- BASTI, G., *Il rapporto mente-corpo nella filosofia e nella scienza*, Bolonia, Studio Domenicano, 1991.
- BEATTIE, J., "Is Putnam Inconsistent and Parochial?", en *Australasian Journal of Philosophy*, LXXI, 1993, pp. 316-321.
- BECHTEL, W., *Filosofía de la mente*, Madrid, Tecnos, trad. esp. de L. M. Valdés, 1991.
- BECHTEL, W., "Currents in Connectionism", en *Mind and Machines*, III, 1993, pp. 125-153.
- BERNIER, P., "Narrow Content, Context of Thought and Asymmetric Dependency", en *Mind and Language*, VIII, 1993, pp. 327-342.
- BERTI, E., "Il concetto di atto nella *Metafisica* di Aristotele", en M. Sánchez Sorondo (ed.) *L'atto aristotelico e le sue ermeneutiche*, Roma, Herder, 1990.
- BLOCK, I., "Order of Aristotle's Psychological Writings", en *American Journal of Philology*, LXXXII, 1961, pp. 50-72.
- BLOCK, N. & FODOR, J., "What Psychological States Are Not", en *Philosophical Review*, LXXXI, 1972, pp. 159-181.



- BLOCK, N. (ed.), *Readings in Philosophy of Psychology*, vol. I, London, Methuen, 1980.
- BLOCK, N., "Functional Role and Truth Conditions", en *Proceedings of the Aristotelian Society*, sup. LXI, 1987, pp. 157-181.
- BLOCK, N., "Can the Mind Change the World?", en George Boolos (ed.) *Meaning and Method. Essays in Honor of Hilary Putnam*, Cambridge, MA, Cambridge University Press, 1990.
- BLOCK, N., "Review of Dennett's *Consciousness Explained*", en *The Journal of Philosophy*, XC, 1993, pp. 181-193.
- BOLOS, G., *Meaning and Method. Essays in Honor of Hilary Putnam*, Cambridge, MA, Cambridge University Press, 1990.
- BREESE, J., *Psicología y vida cotidiana*, Bilbao, Mensajero, trad. esp. de R. López de Uralde, 1973.
- BRESCH, C., "What is Evolution?", en S. Andersen y A. Peacocke (eds.) *Evolution and Creation*, Aarhus, Aarhus University Press, 1987.
- BULLOCK, A. y STALLYBRASS, O. (eds.), *The Fontana Dictionary of Modern Thought*, London, Fontana, 1977.
- BUNGE, M. y ARDILLA, R., *Filosofía de la psicología*, Barcelona, Ariel, trad. esp. de M. A. Galmarini, 1988.
- BUNGE, M., "The Power and Limits of Reduction", en E. Agazzi (ed.), *The Problem of Reductionism in Science*, Dordrecht, Kluwer Academic Publishers, 1991.
- BURGE, T., "Philosophy of Language and Mind: 1950-1990", en *The Philosophical Review*, CI, 1992, pp. 3-51.
- BURNYEAT, M. F., "Aristotelian Philosophy of Mind", en M. C. Nussbaum y A. O. Rorty (eds.) *Essays on Aristotle's De Anima*, Oxford, Clarendon Press, 1992.
- BYNUM, T. W., "A New Look at Aristotle's Theory of Perception", en M. Durrant (ed.), *Aristotle's De Anima in Focus*, London, Routledge, 1993.
- CENCILLO, L., *Hyle*, Madrid, CSIC, 1958.
- CHOZA, J., *Manual de antropología filosófica*, Madrid, Rialp, 1988.
- CHURCHLAND, P. M., "Eliminative Materialism and the Propositional Attitudes", en *The Journal of Philosophy*, LXXVIII, 1981, pp. 67-90.
- CHURCHLAND, P. M., *Matter and Consciousness*, Cambridge, MA, MIT Press, 1988.

- CHURCHLAND, P. M., *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*, Cambridge, MA, MIT Press, 1989.
- COHEN, S. M., "Hylomorphism and Functionalism", en M. C. Nussbaum y A. O. Rorty (eds.) *Essays on Aristotle's De Anima*, Oxford, Clarendon Press, 1992.
- DAVIDSON, D., *Essays on Actions and Events*, Oxford, Oxford University Press, 1980.
- DAVIDSON, D., *De la verdad y de la interpretación*, Barcelona, Gedisa, trad. esp. de G. Filippi, 1990.
- DEL RE, G., "Rapporto sul problema mente-corpo", en G. Del Re y E. Mariani (eds.) *Il rapporto di Napoli sul problema mente-corpo*, Nápoles, IPE, 1991.
- DENNETT, D. C., *Brainstorms: Philosophical Essays on Mind and Psychology*, Montgomery, Bradford Books, 1978.
- DENNETT, D. C., "The Milk of Human Intentionality", en J. L. Garfield (ed.) *Foundations of Cognitive Science*, New York, Paragon House, 1990.
- DENNETT, D. C., *La actitud intencional*, Barcelona, Gedisa, trad. esp. de D. Zadunaisky, 1991.
- DENNETT, D. C., *Consciousness Explained*, London, Penguin Books, 1993.
- DESCARTES, R., *Las pasiones del alma*, Buenos Aires, Aguilar, trad. esp. de C. Berges, 1963.
- DRAY, W. H., "The Historical Explanantion of Action Reconsidered", en S. Hook (ed.) *Philosophy and History*, New York, New York University Press, 1963.
- DRETSKE, F., "Machines and the Mental", en *Proceedings and Addresses of the American Philosophical Association*, LIX, 1985-86, pp. 23-33.
- DURRANT, M. (ed.), *Aristotle's De Anima in Focus*, London, Routledge, 1993.
- EBERT, T., "Aristotle on What is Done in Perceiving", en *Zeitschrift für philosophische Forschung*, XXXVII, 1983, pp. 181-198.
- ECCLES, J. C., *Evolution of the Brain: Creation of the Self*, London, Routledge, 1991.
- ECCLES, J. C., "El misterio del ser humano", en *Nueva Revista*, XXVIII, 1993, pp. 42-65.

- FEIGL, H., *The "Mental" and the "Physical": The Essay and a Postscript*, Minneapolis, University of Minnesota Press, 1958.
- FERRATER MORA, J., *Diccionario de filosofía*, Madrid, Alianza Editorial, 1979.
- FEYERABEND, P., "Mental Events and the Brain", en *The Journal of Philosophy*, LX, 1963, pp. 295-296.
- FIELD, H., "Tarski's Theory of Truth", en *The Journal of Philosophy*, LXIX, 1972, pp. 347-375.
- FODOR, J. A., *Psychological Explanation*, New York, Random House, 1968.
- FODOR, J. A., "Special Sciences (Or: Disunity of science as a working hypothesis)", en *Synthese*, XXVIII, 1974, pp. 97-115.
- FODOR, J. A., *The Language of Thought*, Hassocks, Harvester Press, 1975.
- FODOR, J. A., *The Modularity of Mind*, Cambridge, MA, MIT Press, 1983.
- FODOR, J. A., "Banish DisContent", en J. Butterfield (ed.) *Language, Mind and Logic*, Cambridge, MA, Cambridge University Press, 1986.
- FODOR, J. A., "Why There Still Has to Be a Language of Thought", en William Lycan (ed.) *Mind and Cognition*, Oxford, Blackwell, 1990.
- FODOR, J. A., "Psychosemantics or: Where Do Truth Conditions Come From?", en William Lycan (ed.) *Mind and Cognition*, Oxford, Blackwell, 1990.
- FODOR, J. A., "A Modal Argument for Narrow Content", en *The Journal of Philosophy*, LXXXVIII, 1991, pp. 5-26.
- FODOR, J. A., "Review of K. Sterelny's *The Representational Theory of Mind: An Introduction*", en *Mind and Machines*, III, 1993, pp. 358-362.
- FODOR, J. A., *Psicosemántica*, Madrid, Tecnos, trad. esp. de O. L. González-Castán, 1994.
- GARFIELD, J. L. (ed.), *Foundations of Cognitive Science*, New York, Paragon House, 1990.
- GREGORY, R. L., *The Oxford Companion to the Mind*, Oxford, Oxford University Press, 1987.
- HANNAN, B., "Review of D. C. Dennett's *The Intentional Stance*", en *Mind*, IC, 1990, pp. 291-297.
- HANNAN, B., "Don't Stop Believing: The Case Against Eliminative Materialism", en *Mind & Language*, VIII, 1993, pp. 165-179.

- HARTMAN, E., *Substance, Body and Soul: Aristotelian Investigations*, Princeton, Princeton University Press, 1977.
- HAUGELAND, J., *Artificial Intelligence: The Very Idea*, Cambridge, MA, MIT Press, 1986.
- HAYES, P., HARNARD, S., PERLIS, D. y BLOCK, N., "Virtual Symposium on Virtual Mind", en *Mind and Machines*, II, 1992, pp. 217-238.
- HEMPEL, C. G., "The Logical Analysis of Psychology", en H. Feigl y W. Sellars (eds.) *Readings in Philosophical Analysis*, New York, Appleton Century Crofts, 1949.
- HONDERICH, T., "Smith and the Champion of the Mauve", en *Analysis*, XLIV, 2, 1984, pp. 86-89.
- HORGAN, T. y WOODWARD, J., "Folk Psychology is Here to Stay", en W. G. Lycan (ed.) *Mind and Cognition*, Cambridge, MA, Blackwell, 1990.
- HORGAN, T., "The Austere Ideology of Folk Psychology", en *Mind & Language*, VIII, 1993, pp. 282-297.
- HOUGHTON, D., "Rorty's Talk-About", en A. Malachowski (ed.) *Reading Rorty*, Oxford, Blackwell, 1990.
- HOYNINGEN-HUENE, P., "Theory of Antireductionist Arguments: The Bohr Case Study", en E. Agazzi (ed.) *The Problem of Reductionism in Science*, Dordrecht, Kluwer Academic Publishers, 1991.
- JACKSON, F., "Epiphenomenal Qualia", en *Philosophical Quarterly*, XXXII, 1982, pp. 127-136.
- JAEGER, W., *Aristóteles. Bases para la historia de su desarrollo intelectual*, México, Fondo de Cultura Económica, trad. esp. de J. Gaos, 1946.
- JOHNSON-LAIRD, P. N., *The Computer and the Mind*, Cambridge, MA, Harvard University Press, 1988.
- KHAN, C. H., "Sensation and Consciousness in Aristotle's Psychology", en J. Barnes, M. Schofield y R. Sorabji (eds.) *Articles on Aristotle*, Vol. IV, London, Duckworth, 1979.
- KHAN, C. H., "Aristotle on Thinking", en M. C. Nussbaum y A. O. Rorty (eds.) *Essays on Aristotle's De Anima*, Oxford, Clarendon Press, 1992.
- KIM, J., "Epiphenomenal and Supervenient Causation", en *Midwest Studies in Philosophy*, IX, 1984.
- KRIPKE, S. A., "Naming and Necessity", en D. Davidson y G. Harman (eds.) *Semantics of Natural Language*, Dordrecht, Reidel, 1971, pp. 253-355 y 763-769.



- LEPORE, E. y MCLAUGHLIN, B. P. (eds.), *Actions and Events. Perspectives on the Philosophy of Donald Davidson*, Oxford, Basil Blackwell, 1985.
- LEPORE, E. y LOEWER, B., "Mind Matters", en *The Journal of Philosophy*, LXXXIV, 1987, pp. 630-642.
- LEPORE, E. y VAN GULICK, R. (eds.), *John Searle and his Critics*, Cambridge, MA, Blackwell, 1991.
- LEWIS, D. K., "An Argument for the Identity Theory", en *The Journal of Philosophy*, LXIII, 1966, pp. 17-25.
- LEWIS, D. K., "Psychophysical and Theoretical Identifications", en *Australian Journal of Philosophy*, L, 1972, pp. 249-258.
- LOEWER, B., "From Information to Intentionality", en *Synthese*, LXX, 1987, pp. 307-314.
- LYCAN, W., "Toward a Homuncular Theory of Believing", en *Cognition and Brain Theory*, IV, 1981, pp. 139-159.
- LYCAN, W. G. (ed.), *Mind and Cognition*, Cambridge, MA, Blackwell, 1991.
- LYCAN, W., "A Deductive Argument for the Representational Theory of Thinking", en *Mind & Language*, VIII, 1993, pp. 404-422.
- MCGINN, C., *Mental Content*, Oxford, Blackwell, 1989.
- MCLAUGHLIN, B. P., "Anomalous Monism and the Irreducibility of the Mental", en E. Lepore y B. P. McLaughlin (eds.) *Actions and Events. Perspectives on the Philosophy of Donald Davidson*, Oxford, Blackwell, 1985.
- MCLAUGHLIN, B. P., "What is Wrong with Correlational Psychosemantics", en *Synthese*, LXX, 1987, pp. 271-283.
- MCLAUGHLIN, B. P., "Type Epiphenomenalism, Type Dualism, and the Causal Priority of the Physical", en *Philosophical Perspectives*, III, 1989, pp. 109-135.
- MISKHIN, M., UNGERLEIDER, L. G. y MACKO, K. A., "Object Vision and Spatial Vision: Two Cortical Pathways", en *Trends in Neurosciences*, VI, 1983, pp. 414-417.
- NAGEL, T., "What Is It Like to Be a Bat", en Ned Block (ed.) *Readings in Philosophy of Psychology*, vol. I, London, Methuen, 1980.
- NESBITT, J. y ROSS, J., *Human Inference: Strategies and Shortcoming in Human Judgment*, Englewood Cliffs, NJ, Prentice-Hall, 1980.

- NUSSBAUM, M. C., "Aristotle and Teleological Explanation", en *Aristotle's De Motu Animalium*, Princenton, Princenton University Press, 1978.
- NUSSBAUM, M. C. y PUTNAM, H., "Changing Aristotle's Mind", en M. C. Nussbaum y A. O. Rorty (eds.) *Essays on Aristotle's De Anima*, Oxford, Clarendon Press, 1992.
- NUSSBAUM, M. C. y RORTY, A. O. (eds.), *Essays on Aristotle's De Anima*, Oxford, Clarendon Press, 1992.
- NUYENS, F. J., *L'evolution de la Psychologie d'Aristote*, París-Louvain, Vrin, 1948.
- O'ROURKE, J., "Review of D. C. Dennett's *Consciousness Explained*, of R. Ornstein's *The Evolution of Consciousness: Of Darwin, Freud, and Cranial Fire: The Origins of the Way We Think* and of W. Seager's *Metaphysics of Consciousness*", en *Artificial Intelligence*, LX, 1993, pp. 303-312.
- OWENS, D., "Review of C. McGinn's *Mental Content*", en *Mind*, IC, 1990, pp. 113-122.
- PLACE, U. T., "Is consciousness a brain process?", en *The British Journal of Psychology*, XLVII, 1956, pp. 44-50.
- PLATON, *Phaedrus*, en *Platonis Opera*, vol. II, Oxford, Typographeo Clarendoniano, 1964.
- POLO, L., *Curso de teoría del conocimiento*, vol. II, 2ª Edición, Pamplona, Eunsa, 1989.
- POLO, L., *Quién es el hombre. Un espíritu en el mundo*, Madrid, Rialp, 1991.
- POPPER, K. R. y ECCLES, J. C., *The Self and its Brain*, New York, Springer International, 1977.
- PUTNAM, H., "Minds and Machines", en S. Hook (ed.) *Dimensions of Mind*, New York, New York University Press, 1960.
- PUTNAM, H., "Robots: Machines or Artificially Created Life?", en *Journal of Philosophy*, LXI, 1964, pp.668-691.
- PUTNAM, H., *Mind, Language, and Reality: Philosophical Papers*, vol. II, Cambrige, MA, Cambridge University Press, 1975.
- PUTNAM, H., *Meaning and the Moral Sciences*, London, Routledge & Kegan Paul, 1978.

- PUTNAM, H., *Reason, Truth and History*, Cambridge, MA, Cambridge University Press, 1981.
- PUTNAM, H., *Realism and Reason*, Cambridge, MA, Cambridge University Press, 1983.
- PUTNAM, H., *Representation and Reality*, Cambridge, MA, MIT Press, 1987, trad. esp. de G. Ventureira, Barcelona, Gedisa, 1990.
- QUINE, W. V. O., "Two Dogmas of Empiricism", en *Philosophical Review*, LX, 1951, pp. 20-43.
- QUINE, W. V. O., *Palabra y objeto*, Barcelona, Labor, trad. esp. de M. Sacristán, 1968.
- RAMSEY, W., STICH, S. P. y GARSON, J., "Connectionism, Eliminativism, and the Future of Folk Psychology", en W. Ramsey, S. P. Stich y D. E. Rumelhart (eds.) *Philosophy and Connectionist Theory*, Erlbaum, NJ, Hillsdale, 1991.
- REY, G., "A Reason for Doubting the Existence of Consciousness", en R. Davidson, G. Schwartz, D. Shapiro (eds.) *Consciousness and Self-Regulation*, New York, Plenum, III, 1983, pp. 1-39.
- ROBLES, A., *Hacia la configuración de un campo gnoseológico en el discurso aristotélico sobre el alma*, Madrid, Ed. de la Universidad Complutense de Madrid, Tesis Doctoral, 1982.
- RORTY, A. O., "De Anima: its Agenda and its Recent Interpreters", en M. C. Nussbaum y A. O. Rorty (eds.) *Essays on Aristotle's De Anima*, Oxford, Clarendon Press, 1992.
- RORTY, R., "Mind-Body Identity, Privacy and Categories", en *Review of Metaphysics*, XXIX, 1965, pp. 24-54.
- RORTY, R., "Contemporary Philosophy of Mind", en *Synthese*, LIII, 1982, pp. 323-348.
- RORTY, R., *La filosofía y el espejo de la naturaleza*, Madrid, Cátedra, trad. esp. de J. Fernández Zulaica, 1983.
- ROSS, W. D., *Aristóteles*, Buenos Aires, Charcas, 2ª Ed., trad. esp. de O. N. Guariglia, 1981.
- RYLE, G., *The Concept of Mind*, New York, Barnes and Noble, 1949.
- SCHNEIDER, J., "Omoïoma", en Gerhard Kittel y Gerhard Friedrich (eds.) *Grande Lessico del Nuovo Testamento*, vol. VIII, Brescia, Paideia, trad.

- ital. de F. Ronchi, G. Casanova y benedictinos del monasterio de Santa Escolástica, 1972.
- SEARLE, J. R., "Minds, Brains, and Programs", en *Behavioral and Brain Sciences*, III, 1980, pp. 417-424.
- SEARLE, J. R., *Intencionalidad*, Madrid, Tecnos, trad. esp. de E. Ujaldón Benítez, 1983.
- SEARLE, J. R., "Meaning, Communication, and Representation", en R. Grandy y R. Warner (eds.) *Philosophical Grounds of Rationality, Intentionality, Categories and Ends*, Oxford, Clarendon Press, 1986.
- SEARLE, J. R., "Indeterminacy, Empiricism, and the First Person", en *The Journal of Philosophy*, LXXXIV, 1987, pp. 123-146.
- SEARLE, J. R., *Mentes, cerebros y ciencia*, Madrid, Cátedra, trad. esp. de L. Valdés, 1990.
- SEARLE, J. R., "Response: Meaning, Intentionality, and Speech Acts", en E. Lepore y R. Van Gulick (eds.) *John Searle and His Critics*, Oxford, Blackwell, 1991.
- SEARLE, J. R., *The Rediscovery of the Mind*, Cambridge, MA, MIT Press, 1992.
- SELLARS, W., *Science Perception and Reality*, New York, Routledge & Kegan Paul, 1963.
- SELLARS, W., "Mental Events", en *Philosophical Studies*, XXXIX, 1981, pp. 325-345.
- SLAKEY, T., "Aristotle on Sense Perception", en *The Philosophical Review*, LXX, 1961, pp. 470-484.
- SMART, J. J. C., "Sensations and brain processes", en *The Philosophical Review*, LXVIII, 1959, pp. 141-156.
- SOLMSEN, F., "Greek Philosophy and the Discovery of the Nerves", en *Museum Helveticum*, 1961, pp. 150-197.
- SORABJI, R., "Intentionality and Physiological Processes: Aristotle's Theory of Sense-Perception", en M. C. Nussbaum y A. O. Rorty (eds.) *Essays on Aristotle's De Anima*, Oxford, Clarendon Press, 1992.
- SORABJI, R., "Body and Soul in Aristotle", en M. Durrant (ed.) *Aristotle's De Anima in Focus*, London, Routledge, 1993.
- SOSA, E., "Mind-body Interaction and Supervenient Causation", en *Midwest Studies in Philosophy*, IX, 1984.



- SPERRY, R. W., "Mind, Brain and Humanist Values", en *Bulletin of Atomic Science*, XXII, 1966, pp. 2-6.
- SPERRY, R. W., "Changing Concepts of Consciousness and Free Will", en *Perspectives in Biology and Medicine*, XX, 1976, pp. 9-19.
- STICH, S. P., *From Folk Psychology to Cognitive Science: The Case Against Belief*, Cambridge, MA, MIT Press, 1983.
- TALBOT, M., "Review of H. Putnam's *Representation and Reality*", en *Mind*, XCVIII, 1989, pp. 453-457.
- TARSKI, A., "The Concept of Truth in Formalized Languages", en *Logic, Semantics, Metamathematics*, Oxford, Clarendon Press, 1956.
- TIENSON, J. L., "An Introduction to Connectionism", en J. L. Garfield (ed.) *Foundations of Cognitive Science* New York, Paragon House, 1990.
- TURING, A., "Computing Machinery and Intelligence", en *Mind*, LIX, 1950, pp. 433-460.
- VAN GELDER, T., "Connectionism and the Mind-Body Problem: Exposing the Distinction between Mind and Cognition", en *Artificial Intelligence Review*, VII, 1993, pp. 355-369.
- WILKES, K. V., *Physicalism*, Atlantic Highlands, NJ, Humanities Press, 1978.

## TABLA DE ABREVIATURAS

IA	inteligencia artificial
MBP	<i>mind-body problem</i> (problema mente-cuerpo)
TRM	teoría representacional de la mente





## LOS REDUCCIONISMOS CONTEMPORANEOS ACERCA DE LA MENTE

El estudio de los problemas que entraña una concepción reduccionista de la mente, ha llevado a algunos destacados filósofos de lo mental a modificar sus planteamientos para pasar a defender la tesis de la imposibilidad de una reducción de los fenómenos mentales a fenómenos observables. De entre estos filósofos, he elegido a Hilary Putnam<sup>1</sup> porque fue uno de los primeros defensores del funcionalismo, pero que en los últimos años ha argumentado rigurosamente que la concepción funcionalista de lo mental está equivocada. Estudiaré también a John R. Searle<sup>2</sup> ya que, aunque no puede ser incluido entre los filósofos que han realizado este viraje desde una posición reduccionista a otra no reduccionista, es junto a Putnam otro de los filósofos más representativos entre los que sostienen la tesis de la irreductibilidad de los fenómenos mentales.

### 1. LAS CRITICAS DE HILARY PUTNAM

Putnam afirma que «en mi camino a través de este conflicto, pronto descubrí que la cuestión acerca de la “intencionalidad” ocupaba un lugar central. Que las explicaciones “fisicalistas” del mundo son incompletas, y de forma particular que no tienen en cuenta la intencionalidad, no es, seguramente, una crítica nueva. Brentano, Husserl y otros filósofos hicieron esta crítica, y también la observación de Kant (...) de que no somos capaces de dar una explicación del “esquematismo” en términos de las ciencias naturales, es una versión temprana de la misma denuncia. Pero pocos filósofos actuales esperan que una explicación científica realista, de la intencionalidad tenga la forma de las teorías que Kant, Brentano y Husserl criticaron (es decir, en términos de la “mecánica” del cerebro o en términos de una psicología asociacionista de las “ideas”). Con la aparición de la ciencia de los ordenadores ha nacido un paradigma, totalmente nuevo, de lo que podría ser una explicación realista científica de la intencionalidad. Y esto ha hecho que surja la necesidad, en esta era de los ordenadores, de una exhaustiva investigación del problema de la reductibilidad científica de la intencionalidad»<sup>3</sup>.

En esta investigación acerca de la intencionalidad las críticas de Putnam mostrarán que la naturaleza de la intencionalidad no puede ser explicada reductivamente, ya que es un fenómeno primitivo<sup>4</sup>, irreducible e ineliminable<sup>5</sup>.

### 1.1. La intencionalidad es primitiva

Una característica del reduccionismo de los fenómenos mentales, que es una consecuencia de la afirmación de que los fenómenos mentales pueden ser reducidos a fenómenos observables, es la tesis de que los conceptos son objetos científicamente descriptibles (“psicológicamente reales”), es decir, que los conceptos son entidades que poseemos en nuestra mente o en el cerebro. Por tanto, es normal que a la hora de resolver el problema de la sinonimia, los reduccionistas consideren que lo común a dos términos sinónimos se encuentra en el cerebro. Esta realidad común a ambos términos, y que se encuentra en la cabeza, sería la representación o imagen asociada a los dos términos. Al reconocer que la intencionalidad es primitiva, lo que Putnam pretende es excluir la posibilidad de encontrar alguna propiedad común a todas las ocurrencias de un determinado fenómeno intencional distinta del propio fenómeno<sup>6</sup>. Para ello analiza las teorías representacionistas del mentalismo, ya que si se demuestra que las representaciones no son comunes a todos los seres humanos, entonces no existe un fenómeno primitivo que dé razón de la intencionalidad, y por tanto la intencionalidad ha de ser primitiva puesto que no puede ser primitivizada. A partir de esta propiedad de la intencionalidad es fácil afirmar que no se puede postular la existencia de una *lingua mentis* o lenguaje primitivo.

Putnam caracteriza la tesis representacionista, proveniente de la teoría aristotélica, con los siguientes tres supuestos:

«1°. Toda palabra usada se asocia en la mente del hablante con una representación mental.

2°. Dos palabras son sinónimos (tienen el mismo significado) sólo en el caso de que se asocien con la *misma* representación mental en los hablantes que usan esas palabras.

3°. La representación mental determina a qué se refiere la palabra, en caso de que se refiera a algo»<sup>7</sup>.

Pienso que la atribución a Aristóteles de esta teoría representacionista por parte de Putnam exige un pequeño comentario, ya que considero que no es una tesis aristotélica sino más bien cartesiana. A favor de Putnam tal vez podría proponerse algún texto donde Aristóteles parece sostener esta

posición. Un claro ejemplo es el inicio del libro *De interpretatione*, en el que Aristóteles afirma que:

«Los sonidos vocales son símbolos de las afecciones del alma, y las letras lo son de los sonidos vocales. Y así como la escritura no es la misma para todos, tampoco los sonidos vocales son los mismos. Pero aquello de lo que éstos son primariamente signos, las afecciones del alma, son las mismas para todos, y aquello de lo que estas son imágenes, las cosas reales, son también las mismas»<sup>8</sup>.

La interpretación de este texto ha llevado a filósofos como H. Arens<sup>9</sup> a afirmar que Aristóteles tiene una concepción de los símbolos mentales del tipo de la placa fotográfica. En cambio otros filósofos, entre ellos T. W. Bynum<sup>10</sup>, rechazan que las imágenes aristotélicas sean canónicas. El rechazo de la interpretación representacionista de la teoría aristotélica puede ser iniciado con la consideración de que Aristóteles insiste tanto en el carácter simbólico de la escritura, el habla y el pensamiento, por su rechazo de la teoría de las Ideas de Platón, de forma que puede decirse que sustituye la filosofía de las Ideas platónica por una filosofía del lenguaje. Esta sustitución lleva consigo que Aristóteles considere que el lenguaje es convencional, de forma que la relación de la escritura con la voz, y de la voz con el pensamiento son simbólicas. Por tanto, Aristóteles rechaza que, como proponía Platón, el dialéctico esté llamado a convertirse en el legislador ideal que restablece la correspondencia entre la palabra y la esencia.

La convencionalidad del lenguaje parece implicar que todos los hablantes estarían inmersos en la ebriedad y locura mentales ya que los nombres son dados a las cosas al azar y no hay ninguna relación que justifique el empleo de un nombre determinado. Para resolver este problema lo que hace Aristóteles es añadir una relación de imitación entre el pensamiento y la cosa, con lo que consigue fundamentar esa relación convencional en otra relación de correspondencia más fundamental. Esto hace que Aristóteles hable en este texto de *omoiomata*<sup>11</sup>, pero sería equivocado interpretar este término como identidad<sup>12</sup>. *Omoioma*<sup>13</sup> indica aquello que es hecho igual, es decir, la copia o la imagen. Es sinónimo de *eikon*, y así Platón<sup>14</sup> las emplea para indicar la copia terrena de la realidad original celeste, pero en general es posible establecer una pequeña diferencia ya que *eikon* representaría al objeto, mientras que *omoioma* resalta la semejanza. Por tanto, mientras *eikon* conlleva consigo el ser un icono o arquetipo, es decir, un tipo que sirve de ejemplar y modelo primario al entendimiento y voluntad de los hombres, en cambio *omoioma* indica una forma o semejanza que puede ser causal.



Hechas estas consideraciones, pienso que puede afirmarse, como hace H. Barreau<sup>15</sup>, que aunque Aristóteles defiende una relación de semejanza entre el pensamiento y la cosa, no por ello hay que considerar que esa semejanza implica una copia mental o que considere la mente como una placa fotográfica, y mucho menos que esa copia sea canónica e igual para todos. Aristóteles lo que afirma es que es semejante, pero es equivocado considerar que esa semejanza implica la identidad de la representación mental de los hablantes, la semejanza de la que habla Aristóteles hace relación a la cosa. Aristóteles no puede pensar en una copia mental ya que él habla de afecciones del alma, y para los griegos la afección es el probar, vivir o sentir algo que llega al sujeto desde el exterior, y se opone a aquellos términos que indican una actividad espontánea. Por tanto, es difícilmente sostenible que Aristóteles piense en un tipo de acción que produzca una copia o imagen interna del objeto externo, afirmar tal cosa es ir mucho más allá de lo que dice el texto citado. La imagen de la que habla Aristóteles es creada en el hombre por el objeto externo, y por eso en todo el *De interpretatione* Aristóteles remite continuamente a la realidad externa y nunca a las afecciones, pienso que es muy importante tener esto en cuenta para la interpretación del texto. No pretendo afirmar que Aristóteles no hable de imágenes, incluso reconozco que les da una gran importancia en lo referente al pensamiento<sup>16</sup>, pero hacer depender la sinonimia de la posesión de la misma representación, tal como Putnam interpreta esa expresión, no es aristotélico. Sí que aceptaría que para Aristóteles la sinonimia es consecuencia de las mismas (semejantes) afecciones del alma, pero donde semejante no quiere decir identidad o canonicidad de afecciones, sino similitud que hace referencia a una semejanza en la realidad que produce esas afecciones.

Como lo que me interesa en este momento es señalar los errores de esa tesis representacionista que Putnam define mediante esos tres supuestos, lo que voy a hacer es dejar de lado el problema del origen de esta posición, que pienso que es cartesiano, para centrarme en el estudio de la validez de esos supuestos.

Si se acepta que en la mente existe una misma representación para dos palabras sinónimas, es decir, que la identidad o diferencia de representación es equivalente a la identidad o diferencia de significado, con facilidad se establece la existencia de una *lingua mentis*, de un lenguaje de la mente, que es lo que hace Fodor. Esta concepción considera a la mente como si fuera un criptógrafo, esto es, como un aparato que sirve para cifrar y descifrar un texto del lenguaje de forma enigmática mediante una clave. Este modelo postula que, a un nivel más profundo, la mente posee un lenguaje en el que hay una identidad entre el signo y el significado, y en el que

por tanto cada signo posee un único significado. Dos palabras del lenguaje hablado que tuvieran el mismo significado pero “códigos” distintos, serían la misma palabra en esa *lingua mentis*. Es por tanto un *lenguaje ideal* en el que signos diferentes difieren necesariamente en significado y también en su referencia (no necesariamente en el mundo actual, pero sí al menos en algún mundo posible)<sup>17</sup>.

Putnam argumenta que «lo que hace creíble que la mente (o el cerebro) piense (o “compute”) por medio de representaciones es que todo el pensamiento que conocemos se vale de representaciones. Pero ninguno de los métodos de representación que conocemos —discurso oral, escritura, pintura, grabado en piedra, etc.— tiene la propiedad mágica de *no admitir* representaciones diferentes con el mismo significado. Ninguno de los métodos de representación que conocemos tiene la propiedad de que las representaciones se refieran *intrínsecamente* a lo que habitualmente se refieren, sea ello lo que fuere. Todas las representaciones que conocemos se asocian con su referencia de modo contingente, susceptible de variación a medida que cambia el mundo o la cultura. Este hecho debería bastarnos para desconfiar de las teorías que postulan un reino de “representaciones” con propiedades *tan improbables*»<sup>18</sup>.

Lo que Putnam quiere afirmar es que no existen “representaciones mentales” que simultáneamente satisfagan los tres supuestos de la teoría representacionista. No rechaza, por tanto, la existencia, en un cierto sentido, de representaciones mentales. Lo que no admite es que «sea esencial a la teoría el que la identidad o diferencia de representación sea equivalente a la *identidad de significado*; es decir, que cuando nosotros decimos que dos palabras tienen o no el mismo significado, lo que estamos diciendo es que están o no asociadas a la misma representación mental»<sup>19</sup>.

El mentalismo identifica el significado de una palabra con algo que está en el cerebro/mente de *todo* hablante que sabe usar la palabra, y sostiene que el significado debe ser público. Pero la primera pregunta que es necesario responder para poder hacer responsablemente esta propuesta, es la experiencia que se tiene de que los conocimientos sobre una determinada realidad varían de una persona a otra (p. e., un experto sabe mucho más sobre esa realidad que alguien inexperto), que lleva a pensar que la representación mental o imagen mental que se posee sobre esa realidad debe ser distinta en uno y en otro. En cambio, puede alegar a su favor el fenómeno de que esas dos personas son capaces de hablar de esa realidad y de dar un significado a la palabra con que se la designa. La explicación que Putnam da a estas experiencias se articula en dos partes<sup>20</sup>. En la primera señala que «lo que son las representaciones mentales o las descripciones mentales o las

imágenes mentales en el cerebro/mente de las personas no determina, por lo general, la referencia de una palabra que saben cómo usar»<sup>21</sup>. Esto lo fundamenta en la experiencia de que, aunque la representación mental de uno es distinta de la del otro, ambos se refieren en cambio a la misma realidad. Putnam denomina a este fenómeno como *la división del trabajo lingüístico*, queriendo expresar con ello que «el lenguaje es una forma de actividad cooperativa y no una actividad esencialmente individual. La imagen [representacionista] es errónea cuando supone que todo lo necesario para el uso de la lengua está almacenado en cada mente individual; pero ninguna lengua funciona, en realidad, de esa manera»<sup>22</sup>. «En suma, la referencia se fija *socialmente* y no está determinada por las condiciones u objetos de los cerebros/mentes individuales»<sup>23</sup>.

En la segunda destaca la importancia, para la fijación del significado, del entorno, es decir, de las cosas a las que estamos haciendo referencia. Para ello pone el ejemplo de la palabra “agua”, en la Tierra Gemela, que se emplea para designar a una realidad que no fuera agua, pero que nadie en la Tierra o en la Tierra Gemela es capaz de diferenciarlas. Por tanto se debería decir que la representación mental de los hablantes de la Tierra y de la Tierra Gemela no sería diferente, al menos para el mismo hablante, pero la referencia sí que sería distinta, ya que las sustancias serían diferentes. Putnam llama a este fenómeno «la contribución del entorno». Putnam caracteriza esta contribución diciendo que en toda referencia existe un componente indexical, y así en el caso del agua de la Tierra, este agua vendría caracterizada como la sustancia que se comporta de la misma manera que cualquier otra muestra de agua pura de *nuestro* entorno. Y es este componente indexical el que hace que siendo la misma representación mental, en el caso del agua de la Tierra y de la Tierra Gemela, en cambio su referencia corresponda a dos sustancias distintas. Una vez que esto se demuestra para las sustancias, es fácil trasladarlo a otras realidades naturales. Putnam concluirá diciendo que: «*El entorno desempeña por sí mismo un papel en la determinación de aquello a lo que se refieren las palabras de un hablante, o de una comunidad*»<sup>24</sup>.

De esta forma Putnam sostiene que las propuestas que las teorías reduccionistas hacen en relación al significado y a la referencia se equivocan al no considerar estas dos características. Putnam no pretende dar, con esto, una respuesta a estos problemas, sino que pretende señalar la necesidad de separar las cuestiones relacionadas con la fijación de la referencia de los términos, de las cuestiones acerca de su contenido conceptual. En esta misma dirección Fodor y Chomsky proponen un tipo de representación mental que carece de la tercera propiedad con la que Putnam ha caracteri-



zado la teoría representacionista de la mente (la representación mental determina a qué se refiere la palabra)<sup>25</sup>.

Searle, por su parte, rechaza<sup>26</sup> el segundo presupuesto de la tesis representacionista, el que la identidad de la representación mental dé razón de la sinonimia. Searle rechaza que la sinonimia entre dos términos corresponda a la identidad de la representación mental asociada a ambos, y propone, para evitar esa identidad, añadir a la noción de significado una referencia indexical tanto al sujeto que habla como al tiempo presente<sup>27</sup>. Respecto a esta propuesta lo que Putnam señala es que no resuelve el problema que surge de identificar el significado y la representación mental: «En realidad, uno se ve obligado a recurrir a este tipo de descripciones cuando la palabra no tiene un sinónimo en la lengua que uno habla (...) Pero descripciones tales, como *pájaro de la especie que los natul denominan "chooc"*, no nos dan sinónimos de las palabras cuyo uso se explica de ese modo; más bien, son una manera de soslayar la necesidad de un sinónimo. Una vez más, vemos la imposibilidad de identificar los significados con las descripciones que los hablantes "tienen en la cabeza"; es decir, de identificar las nociones de *significado y representación mental*»<sup>28</sup>.

Siguiendo con los inconvenientes de la tesis de Searle, Putnam considera que las representaciones de un olmo y un haya, si no se sabe diferenciar esas especies de árboles, serán similares para un sujeto inexperto, al mismo tiempo que se es consciente de que debe existir algo que las diferencie, ya que se sabe que un olmo y un haya son realmente distintos. Aparte de la diferente fonética entre un nombre y el otro, no existe ningún otra diferencia entre nuestra representación mental de un olmo y la de un haya, ya que el conocimiento de que existen diferencias entre uno y otro, no hace que las representaciones sean diferentes. Pero en cambio, aunque las representaciones sean las mismas para uno y otro, yo no puedo traducir "olmo" por "haya". De esto Putnam extrae dos nuevas tesis: «primero, que no son *solamente* las "representaciones mentales" las que se preservan en la traducción y, segundo, que las "representaciones mentales" no bastan para fijar la referencia»<sup>29</sup>.

La propuesta que, para salvar su teoría representacionista, Fodor hizo aunque luego no la publicara<sup>30</sup>, fue la de hablar de la noción de "contenido estricto" (*narrow content*) que sería la representación mental a su más bajo nivel (la "representación semántica" en "Mentalese"), y de "contenido amplio" (*broad content*) que sería otro referente y que nos daría el(los) referente(s) para los mundos posibles. Este contenido estricto es una función de las propiedades observables de varias cosas, más en concreto, lo que Fodor propone es que el cerebro contiene dispositivos para reconocer

*pautas o patrones (patterns)*. Por tanto, en algún lugar de mi cerebro lo que hago es construir una subrutina que tiene la función de reconocer cosas que presentan esa pauta particular. A esta propuesta responde Putnam indicando que «es esencial a la afirmación básica de Fodor (la sistemática ambigüedad de la noción de *significado*) que *todo* término del lenguaje tenga un “contenido estricto” (*narrow content*) (...) Ciertamente, Fodor ha argumentado que es imposible dar cuenta de la *creencia* a menos que cada cosa que pueda ser creída tenga un contenido estricto. Pero la palabra “*zeitgeist*”, por ejemplo, no está asociada con un prototipo perceptual, aunque podamos tener creencias sobre la misma»<sup>31</sup>. Una segunda dificultad vendría del hecho de considerar el contenido estricto como una función de las propiedades observables del referente, ya que esto presenta dificultades de dos tipos: «el prototipo perceptual no es preservado en la traducción»<sup>32</sup>, esto se ve cuando se traducen términos de dos culturas diferentes, «y, en algunos casos, no es lo que importa para la traducción»<sup>33</sup>, por ejemplo en el caso del término “bruja”, lo que importa no es que sea una mujer fea y mayor, con una gran nariz; sino que lo que importa son sus poderes mágicos.

Fodor sí que ha publicado, en cambio una concepción del contenido estricto<sup>34</sup> con la que intenta dar respuesta al problema de la Tierra Gemela, que, aunque es un poco distinta respecto a la que Putnam acaba de examinar<sup>35</sup>, sigue sin dar respuesta a las objeciones planteadas. Su tesis es que el contenido estricto de mi pensamiento, cuando creo que los lagos contienen agua, es una función que toma mi contexto actual y lo pone en la proposición que afirma que los lagos contienen agua, mientras que en la Tierra Gemela toma el contexto actual de mi gemelo y lo pone en la proposición que afirma que los lagos contienen XYZ. Defino el contenido estricto como función ya que Fodor huye de definirlo como contenido, ya que si intento enunciarlo deja de ser contenido restringido para pasar a ser contenido amplio. «El contenido restringido es, de una forma radical, inexpresable, porque sólo es contenido *de una forma potencial*»<sup>36</sup>.

Lo que propone Fodor es que el agua y el XYZ son tan similares, que esa similitud se debe manifestar también en los pensamientos. Por tanto pueden ser caracterizados como pensamientos del mismo tipo ya que esa generalización también me permite explicar la similitud de sus comportamientos. Dada la identidad neurofisiológica que postula el ejemplo, si «yo estoy en el contexto de mi Gemelo, mis pensamientos acerca del «agua» son acerca de XYZ, si y sólo si los suyos lo son»<sup>37</sup>. «Específicamente, tu pensamiento es idéntico en contenido al mío sólo si en todo contexto en el que tu pensamiento tiene la condición de verdad T, el mío tiene la condición de verdad T, y viceversa»<sup>38</sup>. La objeción que se descubre inmediatamente a

esta argumentación está reconocida por Fodor al aceptar, en un artículo posterior<sup>39</sup>, que al decir yo en la Tierra Gemela que 'quiero agua' y al decirlo mi gemelo en la Tierra, ninguno de los dos conseguimos lo que hemos pedido, y que, «por tanto, nuestros comportamientos continúan siendo relevantemente diferentes bajo estas descripciones intencionales»<sup>40</sup>. Es decir, cuando yo me traslado de la Tierra a la Tierra Gemela, sigo llevando conmigo mi contexto, y los pensamientos que tengo son de agua y no de XYZ. Aun reconociendo esto, Fodor terminará afirmando que «los pensamientos acerca del tagua<sup>41</sup> y los pensamientos acerca del agua no tienen poderes causales diferentes. Por tanto, para los propósitos de los psicólogos, son el mismo estado intencional. Pero no pueden ser el mismo estado intencional a menos que tengan el mismo contenido intencional. Y no pueden tener el mismo contenido intencional a no ser que el contenido intencional sea individuado estrictamente»<sup>42</sup>.

Para compatibilizar estas dos afirmaciones lo que hace Fodor es defender que «ser un pensador de agua es el mismo poder causal que ser un pensador de tagua, sólo que instanciados en una persona con una historia causal diferente»<sup>43</sup>. Fodor considera que en los dos casos el poder causal es idéntico, ya que: «a) si su proferencia (/pensamiento), hubiera ocurrido en mi contexto, *habría tenido* los efectos que tuvo mi proferencia (/pensamiento), y b) si mi proferencia (/pensamiento) hubiera ocurrido en su contexto, *habría tenido* los efectos que tuvo su proferencia (/pensamiento)»<sup>44</sup>, por ello afirma que para lograr esa identidad esta «tiene que ser fijada a lo largo de los contextos y no dentro de los contextos»<sup>45</sup>, ya que las propiedades relacionales de los estados psicológicos son irrelevantes para sus poderes causales. Pero Fodor necesita fundamentar esta afirmación mediante una teoría del contenido amplio en la cual pueda dotar de sentido a la noción de historia causal que relaciona el tipo mental y el contexto en el cual el tipo mental es construido, pero no parece haberlo conseguido<sup>46</sup>, y no lo ha conseguido porque no ha sido capaz de dar razón de la determinación del contenido restringido, porque mientras que no pueda usarse el contenido restringido para explicar el contenido amplio, esa función que relaciona el tipo mental y el contexto no se puede determinar.

La teoría de Fodor no logra resolver las dificultades que Putnam advertía. En primer lugar, sigue siendo cierto que a la palabra "*zeitgeist*", difícilmente puede poseer un contenido estricto, tal vez pudiéramos pensar que lo sea su fonética, pero el mismo Fodor rechaza tal posibilidad al considerar que el contenido estricto es un contenido de una forma potencial. Pero sorprendentemente, como señala Putnam, yo puedo tener creencias acerca de ese término. La otra dificultad es la de la traducción, que aunque



pueda parecer que queda resuelta al carecer el contenido estricto de un contenido preciso, parece que más bien ocurre todo lo contrario, ya que la consideración del contenido estricto como una función del contexto al referente lleva consigo el que la función esté en la cabeza de la persona. Esta posición posee los mismos inconvenientes que tenía la de Searle ya que identifica el significado con algo que está dentro de la cabeza de la persona.

Block ha propuesto una tesis un poco distinta<sup>47</sup>. El contenido estricto corresponde a lo que el hablante tiene en la cabeza, pero no como función de las propiedades observables, sino según su papel conceptual (*conceptual role*). El papel conceptual está relacionado con qué tipo de creencias contiene esa palabra, con las inferencias que se obtienen de ella, con el tipo de conductas con las que ésta se relaciona, etc. Para Block, «las palabras son similares en contenido estricto cuando tienen “papeles conceptuales” semejantes (...) en sus respectivos lenguajes. Y cuando las palabras son semejantes en el contenido estricto y, además, tienen la misma referencia (o mejor, el mismo contenido amplio), entonces son semejantes en el significado»<sup>48</sup>.

Putnam señala tres dificultades en esta teoría. «(1) el papel conceptual sólo funciona significativamente para fijar el significado de ciertas clases de palabras, y no de otras» (p. e., los términos que designan cosas naturales —agua, vegetación— cuyo significado se relaciona más con su extensión). «(2) aunque el papel conceptual desempeña una función importante en la descripción del significado, no es el *único* que lo describe»<sup>49</sup> (p. e., la palabra bruja parece que forma parte de una creencia muy estable, que es la de ser una mujer que tiene poderes mágicos que le han sido dados mediante un pacto con el diablo, pero podemos encontrar brujas buenas o brujas que no han pactado con el diablo ya que en su cultura el diablo es desconocido). Por último, Putnam señala una tercera dificultad en el concepto técnico que Block tiene del papel conceptual, ya que si los dos problemas anteriores se dan en un concepto más amplio del papel conceptual, con mucha más razón se darán en el concepto que Block defiende.

J. Beattie afirma que toda esta propuesta de Putnam es inconsistente<sup>50</sup>, aunque, al mismo tiempo, considera que Putnam logra mostrar mediante estos razonamientos los grandes inconvenientes que posee la teoría representacionista. Para mostrar esa inconsistencia lo que hace es suponer que lo que Putnam propone son dos teorías acerca de la fijación de la extensión de un término, una que se basa en el entorno y otra en la comunidad social. Si la propuesta de Putnam fuera ésta, entonces surge el problema de hacer que cooperen entre sí esas dos teorías en el caso de que la comunidad social se equivocara al fijar la extensión de un término, ya que

entonces la fijación de la extensión de ese término difiere entre una teoría y la otra. Pero es que no es esto lo que Putnam pretende. Lo que Putnam busca es demostrar que, en la fijación del significado de un término, debe tenerse en cuenta el entorno y la comunidad social, ya que el significado de los términos no puede reducirse a la representación que se encuentra en la mente del hablante. No propone una nueva teoría, sino que muestra que además de la representación es necesario considerar otros factores a la hora de fijar el significado de un término.

Otros filósofos han comenzado a proponer, al considerar el argumento de la Tierra Gemela, lo que McGinn ha llamado 'externalismo', e incluso se empieza a diferenciar un externalismo fuerte y otro débil. «Dicho crudamente, el externalismo fuerte liga los estados mentales con una *parte* particular del mundo -la parte en la que el mismo sujeto habita- mientras que el externalismo débil liga los estados mentales al mundo en su totalidad, sin exigir que esa entidad no mental sea el lugar donde el sujeto se encuentra localizado»<sup>51</sup>. El problema que McGinn ve en ese externalismo es presuponer que aceptar esta teoría conlleva consigo rechazar, por su incompatibilidad, la consideración de que la mente sea una substancia. Su argumentación<sup>52</sup> se inicia con la consideración de que la individuación de los estados mentales depende de propiedades y objetos que están fuera de nuestra mente, y por tanto, que la mente tendrá, como elementos que la constituyan, esos objetos. Si no fuera así, habría que dejar de pensar que la mente es algo que ocupa una determinada posición en el espacio y en el tiempo, ya que no encuentra forma de armonizar esa dependencia respecto a lo externo y la substancialidad de la mente, si no es por medio de la existencia en mi mente de esos objetos. Como es absurdo que los objetos externos formen parte de mi mente, entonces mi mente no es una substancia.

Este argumento confunde esa dependencia del significado respecto al entorno en el que la persona se encuentra, con la necesidad de que ese entorno forme parte de la mente. Putnam nunca ha pretendido afirmar tal cosa, a este respecto es interesante el ejemplo que propone David Owens: «El Príncipe Carlos es un Windsor en virtud del hecho de que su madre es una Windsor (por tanto su ser un Windsor está, en parte, constituido por ser su madre una Windsor), pero la madre del Príncipe Carlos no es parte del Príncipe Carlos, y tampoco lo es la casa de los Windsor»<sup>53</sup>. Putnam lo que propone es que al querer fijar el significado de un término, es necesario considerar el entorno, pero no como una parte de la mente, sino como un elemento más que determina el significado del término que empleo.

McGinn señala, en cambio, una dificultad que parece favorecer más una visión internalista que una externalista. Esta dificultad surge al considerar el caso de la Tierra Gemela pero en la que mi Gemelo no tiene músculos en el antebrazo. ¿Sentiría él la misma sensación que siento yo cuando me duele el antebrazo? En un principio, la intuición más fuerte es la de aceptar que mi Gemelo no podrá tener tal sensación, y que por lo tanto hay que aceptar un cierto externalismo. Pero también es posible dar otra solución, que es la de considerar el cuerpo como interno. Enfrenta, de esta forma, una postura internalista que defendería que los estados del cuerpo de una persona determinan la identidad de sus estados mentales, frente a un externalismo que negaría esto, y termina inclinando la balanza por el lado del internalismo, ya que el dolor en el antebrazo depende de que se tenga antebrazo. Considero que es interesante revalorizar el valor del cuerpo en la explicación de la realidad de la mente, pero pienso que este externalismo poco tiene que ver con lo que Putnam propone, ya que en ningún momento niega la importancia del cuerpo en la individuación de los estados mentales, lo que señala es que también es importante, en esa individuación, el entorno que lo rodea.

## 1.2. La intencionalidad no es reducible

La invención y consecuente aparición de los ordenadores generó la esperanza de que los modelos computacionales podrían explicar los fenómenos mentales, y así se confiaba en dar una explicación satisfactoria de la intencionalidad. Pero poco a poco se ha ido viendo que tal explicación reductiva no es tan sencilla, y ha comenzado a considerarse que tal vez no sea factible tal reducción.

### 1.2.1. *El significado no es reducible*

Putnam analiza los problemas del significado en los tres primeros capítulos de *Representation and Reality*, estableciendo una estrecha conexión entre los problemas del significado y la fijación de creencias, y mostrando que es una equivocación pensar en el significado como un objeto que puede ser aislado científicamente<sup>54</sup>. El objetivo de su argumentación es poder afirmar la irreducibilidad de la intencionalidad basado en la irreducibilidad del significado.

Putnam comienza esta argumentación aportando las tres razones por las que —a su juicio— estas teorías se equivocan<sup>55</sup>:



— *El significado es holístico*, como ya propuso Quine<sup>56</sup> en oposición a la tesis del positivismo lógico que sostenía que todo término puede ser definido mediante un número limitado de términos llamados observacionales. En cambio el holismo del significado afirma que lo que tiene valor es el conjunto de las frases, y ese valor no es la suma del valor experiencial de las frases individuales: El lenguaje describe la experiencia como si fuera una red. Una consecuencia del holismo del significado es que es falso considerar que la definición explica o fija el significado de las palabras, ya que «la mayoría de los términos *no pueden* ser definidos —si por “definición” se entiende algo que se fija de una vez y para siempre, algo que aprehende absolutamente el significado del término»<sup>57</sup>.

— *El significado es en parte una noción normativa*. Por ello toda interpretación depende de la benevolencia o beneficio de la duda, ya que siempre se debe tener presente la posibilidad de tener que descartar al menos algunas diferencias de creencias cuando interpretamos. Esto hace que al interpretar una proposición, se trate a los conceptos como si tuvieran «una identidad a través del tiempo, pero no esencia»<sup>58</sup>. Esto hace que la decisión de interpretar a alguien, de una determinada forma o de otra, esté ligado íntimamente a juicios normativos.

— *«Nuestros conceptos dependen de nuestro entorno físico y social de una forma que la evolución (que en lo que respecta a nuestros cerebros, terminó hace unos 30.000 años) no podía prever.*

Para suministrarlos un repertorio innato de nociones que incluye carburador, burócrata, energía potencial, etc., como lo exige la versión fodoriana de la Hipótesis de lo innato, la evolución tendría que haber sido capaz de anticipar todas las contingencias de los entornos físicos y culturales futuros. Obviamente, no lo hizo ni lo podía hacer»<sup>59</sup>.

Pero la refutación de las tesis reductivistas del significado sólo se consigue cuando estas tres razones actúan juntas, ya que cada una por separado puede ser soslayada si no se tiene en cuenta a las otras dos<sup>60</sup>. Con el holismo del significado se afirma la imposibilidad de definir, como pretendía el positivismo, todos nuestros conceptos a partir de un stock de nociones primitivas, pero al mismo tiempo esta argumentación en contra del positivismo se apoya en la no violación de que el significado es en parte una noción normativa, ya que se debe fundamentar el holismo en un concepto de “significado” que permita considerar que, aunque “cambie el significado” en elementos relevantes del término, este se mantiene idéntico por medio del proceso de justificación y fijación de creencias. Es decir, que descubrir por ejemplo que el momento no es exactamente el producto de la

masa por la velocidad, no quiere decir que el término momento haya cambiado de significado, sino que al interpretarlo se tendrá presente esa diferencia de creencia. Pero no por ello se afirmará que sea un nuevo término porque haya variado su descripción, sino que se sigue hablando del mismo "momento", y por tanto se seguirá empleando para designar el concepto antiguo y el moderno, pero eso sí, con una diferencia de creencia pero no de significado.

Es importante unir estas dos razones, ya que si sólo se considera el holismo, inmediatamente las teorías reduccionistas pedirían que se dieran razones que justifiquen el que consigamos entendernos. El que se pueda tener una conversación con otra persona ha llevado a los reduccionistas a buscar siempre un algo común en los términos que ambos interlocutores emplean. Lo que Putnam pretende mostrar es que ese algo común no puede encontrarse en los mismos términos, y eso se logra con el holismo, pero es que entonces puedo pensar en la oración como una estructura en la que además de los términos que la forman, existen unas reglas que añaden nuevos significados. Esto hace necesario considerar el carácter normativo del significado que conlleva el que esa oración sea interpretada descartando diferencias de creencia, y ese beneficio de la duda no puede ser formalizado.

También ocurre algo similar al considerar la posibilidad de que sea el proceso evolutivo el que genere en mí el conjunto de términos que poseo. Esta tesis resulta muy atractiva si se pretende resolver el problema de la dependencia del contenido de mis representaciones al entorno que me rodea, ya que estos términos no dependerían del entorno, y por lo tanto sobre ellos sí que puedo basar una teoría de la representación que sea solipista. El inconveniente de esta tesis es dar razón del proceso evolutivo que ha dado lugar a los términos que empleo, ya que si pretendo que todos los términos se han generado evolutivamente, entonces la evolución tendría que haber sido capaz de anticipar todas las contingencias de los entornos físicos y culturales futuros, y esta tesis es realmente difícil de defender. Es necesario encontrar una teoría que haga esta dependencia entre el término y el proceso evolutivo, al menos, defendible.

Una solución posible es considerar que lo que se logra mediante el proceso evolutivo es un stock de nociones básicas, a partir de las cuales se definen las demás, y por tanto no sería necesario que la evolución anticipara todas las contingencias de los futuros entornos físicos y sociales. Pero entonces es cuando entra en juego el argumento del holismo del significado ya que permite afirmar que los términos no pueden ser definidos a partir de un conjunto de términos más primitivos y reducidos, sino sólo mediante el léxico completo de ese lenguaje. Que sea necesario el considerar



todo el léxico del lenguaje se basa en la afirmación de que la definición no fija de una vez y para siempre el significado del término. Por lo tanto, si el significado de un término puede cambiar, la evolución debe anticipar, necesariamente, todos los cambios de los términos del lenguaje. Una posible solución es considerar que esos cambios se producen evolutivamente en ese stock básico de términos, pero entonces el holismo tendría que ser falso, ya que las definiciones de los términos no primitivos sí que podrían ser fijadas de una vez y para siempre mediante ese stock básico.

### 1.2.2. *El funcionalismo no funciona*

Putnam rechazó el funcionalismo que él propuso inicialmente por considerar que la definición en términos de una máquina de Turing de los estados mentales no daba una adecuada representación de la psicología del hombre y de los animales. Pero basándose en la contribución del entorno, es decir, en que no podemos individuar conceptos y creencias sin hacer referencia al entorno, pretende ir aún más lejos y mostrar que los modelos computacionales del cerebro/mente no son suficientes para la psicología cognitiva, ya que el fenómeno que estos modelos buscan para explicar la intencionalidad —el “proceso mental”— no existe.

Se podría pensar que una forma de solucionar esta referencia al entorno es mediante la inclusión de aspectos del entorno en la noción de estado computacional. Esto es lo que hace el sociofuncionalismo al caracterizar la referencia como una relación funcional entre las representaciones usadas por los organismos y cosas que pueden estar dentro o fuera de esos organismos<sup>61</sup>. «Este punto de vista sería razonable si mi objeción al funcionalismo se basará únicamente en el hecho de que la determinación del significado depende en parte del entorno social y en parte de la naturaleza de las cosas externas»<sup>62</sup>. Pero «*el significado y la referencia dependen de lo que denominé “descontar las diferencias de creencia”*»<sup>63</sup>.

Cuando se interpreta lo que alguien dice, simultáneamente se adscriben creencias y deseos a esa persona. Pero esa adscripción nunca podrá lograr que esas creencias y deseos sean los suyos. Por tanto, si se pretende afirmar la sinonimia entre una palabra que yo empleo y la que emplea otra persona, necesito aplicar un criterio de “razonabilidad” respecto a las creencias que atribuyo al otro. Este criterio de “razonabilidad” lleva consigo el dejar en un segundo lugar las creencias que yo pueda tener de esa realidad y de su entorno, para intentar lograr hacer razonable la imagen que intenta transmitirme la otra persona sobre esa realidad. Por tanto, aunque en el futuro «una definición funcionalista de la sinonimia y la correferencialidad for-

*malizara (y, probablemente, “reconstruyera racionalmente”) estos criterios intuitivos de razonabilidad (...). La idea de construir hoy en día tal definición de sinonimia y correferencialidad es totalmente utópica»<sup>64</sup>.*

Pero aunque no se pueda hacer en la práctica, eso no evita que se pueda considerar que es posible hacerlo “en principio”. Es decir, aunque todavía no se haya podido lograr una teoría que defina la sinonimia y la correferencialidad mediante términos funcionales, no implica que el logro de tal teoría sea imposible. Esto supone afirmar que es posible, en principio, definir la referencia en términos físico/computacionales, de forma que su extensión en cualquiera de los mundos posibles coincida con la de “referirse a” en esos mundos posibles<sup>65</sup>. Pero el problema de esta teoría es que no es capaz de extraer, de su propio modelo, una noción de sinonimia que permita señalar el hecho de que cuando dos personas pronuncian la misma palabra lo hacen con el mismo significado<sup>66</sup>. No es posible lograr este concepto de sinonimia ya que por muchas funciones que yo proponga en mi sistema informático, al menos alguna de ellas tendrá que depender de la experiencia adquirida por el propio sistema informático, por lo tanto esta función hará que el estado interno del sistema sea distinto incluso para una misma palabra. Una forma de resolver este problema es considerar que esos términos pueden construirse analíticamente, es decir, mediante una definición, pero con esto sólo se logra pasar el problema a los términos con los que se hace esa definición, ya que si son términos adquiridos vuelve a producirse el mismo problema. Sólo «si hubiera un repertorio de palabras universales y biológicamente innatas —digamos, “términos observacionales”— en función del cual todas las palabras fueran analíticamente definibles, entonces la definición de “analiticidad” y la identificación de estos conceptos básicos lingüísticamente universales resolverían el problema para el caso de los seres humanos»<sup>67</sup>. Pero como se ha visto, el holismo del significado imposibilita que esto pueda realizarse, ya que el significado de un término también depende de la oración como un todo, y los términos por sí solos no dan razón de esta estructura.

Por lo que respecta a que un término observacional pueda ser innato, es conveniente señalar el ejemplo del “*gavagai*” de Quine<sup>68</sup>, que muestra cómo la identidad en los “significados analíticos” no es condición necesaria para la sinonimia, incluso en los casos de términos observacionales. Quine plantea el caso de la traducción, sin poder contar con la ayuda de un intérprete, del lenguaje de un pueblo al que se llega por primera vez. Supone que estudiando el lenguaje de un indígena pasa por delante de los dos un conejo y el indígena dice ‘*gavagai*’, entonces el lingüista anota ‘conejo’ como traducción provisional. Posteriormente el lingüista tendrá que ir com-

probando sus conjeturas mediante la combinación de sentencias indígenas con situaciones-estímulo para observar la aprobación, recusación o inhibición del indígena. Pero también aquí surge el problema de la necesidad de conjeturar las actitudes del indígena. Al considerar todas estas dificultades, Quine propone entender el agente que provoca el asentimiento del indígena a '¿Gavagai?' como estimulaciones, y no como conejos, ya que el indígena puede asentir a 'gavagai' por un movimiento mal percibido de la hierba, ya que su asentimiento depende de una información paralela anterior. Por ejemplo, si suponemos que en ese país existe una mosca del conejo, que el lingüista desconoce, pero que el indígena reconoce, entonces un animal mal percibido pero cerca del cual hay moscas de ese clase, podría ser reconocido por el indígena como 'gavagai'. La consecuencia que deduce Quine de toda esta argumentación es que, aunque pueda aislarse algo que es común y que prescindir de la información lateral o paralela, de forma que se consiguiera una significación estimulativa más estable y más social, no por ello se salva la dependencia de ese denominador común a la información socialmente compartida. «No hay ningún criterio evidente de la eliminación de tales efectos para dejar sólo la significación propiamente dicha de 'Gavagai', en el supuesto de que eso de significación propiamente dicha sea algo»<sup>69</sup>. Por tanto, «cuando se traduce 'Gavagai' por 'Conejo' a pesar de las discrepancias entre significaciones estimulativas antes imaginadas, lo hace porque dichas significaciones estimulativas parecen coincidir en grado predominante, mientras que las discrepancias, en la medida en que consigue identificarlas, parecen susceptibles de eliminación, o despreciables al menos, como efectos de interferencias imprecisas»<sup>70</sup>.

Se debe rechazar por tanto la idea de que los estados mentales, de los hablantes que tienen una misma referencia, sean idénticos; para pasar a considerar que los estados son «*equivalentes respecto de alguna relación de equivalencia que, a su vez, sea computable, o al menos definible en el lenguaje de la teoría computacional más el lenguaje de la física*»<sup>71</sup>. Lo que ocurre es que, pensar que la sinonimia es, desde un punto de vista teórico, idéntica con una relación computable, conlleva dos clases de problemas.

El primero proviene de que «es falsa la hipótesis de que es posible, en principio, decir a qué se refiere un término usado en un entorno a partir de una descripción suficientemente completa del entorno en términos de algún conjunto estandarizado de parámetros físicos y computacionales, *a menos que amplíemos la noción de entorno del hablante de modo que incluya todo el universo físico*»<sup>72</sup>. No pretende Putnam negar con esto que no se puedan lograr formalizaciones de nuestros entornos que permitan a un ordenador ser capaz de manejarse con una cierta fiabilidad en ese entorno, pero



sí que quiere mostrar que la pretensión de lograr formalizar el entorno siempre dejará de lado aspectos que después influirán en el significado de alguno de los términos.

El segundo proviene de que «cualquier teoría que “defina” la correferencialidad o sinonimia tiene que examinar, de algún modo, todas las teorías posibles (...) Ciertamente, un algoritmo así no se limitaría a “simular” una facultad que existe, efectivamente, en los seres humanos, pues ningún ser humano puede revisar toda la matemática posible, toda la ciencia fáctica posible, etc»<sup>73</sup>. Tal algoritmo «sólo un dios podría escribirlo, o un ser humano mucho más inteligente que los seres humanos de todas las posibles sociedades humanas que pudiera examinar todos los modos humanos posibles de razonamiento y conceptualización»<sup>74</sup>. La necesidad de examinar todas las teorías posibles es una consecuencia del propio concepto de sinonimia, ya que hay veces en que dos términos tienen el mismo significado porque existe una teoría científica que muestra su identidad. Por tanto, toda teoría de la sinonimia y la correferencialidad debe dar razón de estas identidades.

Pero que no se pueda escribir tal algoritmo no impide que, en principio, exista. Es decir, un problema es que esa teoría de la sinonimia se convierta en una teoría muy compleja y otra cosa muy distinta es que tal teoría no se pueda lograr. Como el número de todas las posibles sociedades humanas y de todos los modos humanos posibles de razonamiento y conceptualización, no son infinitos, entonces parece posible lograr tal formalización. Pero es que «la idea de buscar una relación de equivalencia computable (o incluso bien definida) entre los estados funcionales que corresponden a la relación de equivalencia implícitamente definida por la práctica de los “buenos intérpretes” vuelve a tropezar con una dificultad insuperable en la infinitud abierta e inagotable de los esquemas conceptuales que, “en principio”, han de ser interpretados»<sup>75</sup>. Este problema, con el que debe enfrentarse toda teoría que pretenda una correcta interpretación de las actitudes proposicionales, hace que la función lógica compleja que es necesaria para tal interpretación sea del tipo de una “lista infinita”. Por tanto, «la afirmación de que existe algo finitamente especificable, común a todos los casos de interpretación correcta, es una afirmación en la que no creemos, pues no hay razones que la justifiquen. Y si la asignación de una actitud proposicional depende de la interpretación (...), entonces, que la relación de equivalencia se defina por una “lista infinita” (...) significa que el “estado” correspondiente a cualquier actitud proposicional (...) será igualmente definido por una lista infinita que no está construida por ninguna regla»<sup>76</sup>. Por tanto, lo que lograríamos “en principio” no es una reducción del fenómeno,

sino el reconocimiento de que ese fenómeno sólo es reductible a sí mismo, es decir, que es irreductible.

«La única opción realista que le queda abierta al funcionalismo es la de enunciar su tesis de modo que no dependa del formalismo computacional»<sup>77</sup>. Esto es lo que propone David Lewis con su teoría de la “psicología popular” (*folk psychology*). Esta se apoya en dos nociones: la de “estado neuroquímico” y la de “papel causal”. Para Lewis cada estado mental es idéntico con un estado neuroquímico, pero no con un estado que sea idéntico para diferentes especies. Pero en cambio, aún en el caso de especies diferentes, estos estados tienen el mismo papel causal. Por tanto, «si es posible hallar un conjunto ordenado de “estados físicos” que desempeñen los “papeles causales” especificados por una teoría (...), entonces se dice que ese conjunto de estados es una “realización” de la teoría. Lo que Lewis afirma (...) es que cualquier realización de la psicología popular es una interpretación pretendida de la psicología popular»<sup>78</sup>. Pero «la exigencia de que una teoría tenga una realización es poco sólida y no sirve a los propósitos de Lewis, si se admite que las “realizaciones” pueden implicar estados físicos arbitrarios»<sup>79</sup>, ya que significaría que la teoría psicológica que propone tendría excesivas realizaciones, por tanto es necesario restringir el tipo de realizadores posibles a la disyunción de los estados físicos básicos. Además este funcionalismo implicaría un conductismo, ya que «si es cierto que poseer determinados estados mentales equivale simplemente a poseer cierta “organización funcional”, ¿entonces también es cierto que poseer determinados estados mentales es simplemente poseer ciertas disposiciones de conducta!»<sup>80</sup>. Pero es que el concepto de causalidad que Lewis emplea no es el de la física matemática, sino que se basa «en la “similitud de mundos posibles” y en la “teoría de los acontecimientos”, que, a su vez, presupone su “teoría de los universales”»<sup>81</sup>, de forma que lo que Lewis propone es que los términos mentales lo que especifican es un nexo de agentes causales que utilizamos para explicar la conducta de los agentes cognitivos. Pero es que con todos estos nuevos términos «la hipótesis de Lewis *no* consiste en una reducción de las actitudes proposicionales a algo físico, sino, más bien, en una reducción de las actitudes proposicionales a un conjunto de propiedades y relaciones altamente metafísicas»<sup>82</sup>. Al mismo tiempo, esta reducción significa que «su teoría requiere que cada organismo al que se aplica la psicología popular será un modelo para dicha psicología (...), es decir, que cada organismo posea un estado físico para cada actitud proposicional,»<sup>83</sup> y ya se ha visto que tal idea es falsa. Pero es que, aún admitiendo que existiera un estado físico por actitud proposicional, la teoría de Lewis no consigue determinar tales estados de manera suficiente mediante las “trivialidades”<sup>84</sup> de la psicología popular, ya que «si el cuento de las trivialidades



fuera lo único que hay con respecto a la interpretación, nadie podría adquirir ningún concepto que la cultura no tuviera ya (o que no fuera reductible a los conceptos que posee esa cultura). Nuestras trivialidades no son el fundamento al que pueden reducirse todas las nociones posibles»<sup>85</sup>.

Putnam concluirá los capítulos dedicados al funcionalismo diciendo:

«El funcionalismo comenzó por rechazar la idea ingenua de que las descripciones de actitudes proposicionales (por ejemplo, “creer que hay muchos gatos en el vecindario”) corresponden de manera biunívoca a estados del cerebro, independientemente de la especie. Sin embargo —y éste es su talón de Aquiles— da por sentado que corresponden biunívocamente a estados del cerebro en cada *organismo individual*. Pero esto es sólo “ciencia popular” en un sentido peyorativo. Los seres humanos difieren enormemente entre sí y de una a otra cultura (...). La verdadera práctica de la interpretación es de final abierto y prácticamente extendible hasta el infinito (...). Hemos considerado la respuesta de que un ser “más inteligente” que el *homo sapiens* podría, “en principio”, (...) construir una clase de equivalencias de estados computacionales en la cual debe hallarse todo ser humano para tener una creencia particular. Pero entonces surgió el mismo problema al preguntar *qué significa para ese ser más inteligente tener una creencia* (...). Finalmente, hemos considerado la idea de permitir que las actitudes proposicionales sean “implícitamente definidas” por las teorías. Pero ignoramos qué clase de teoría podría hacerlo (aun si su alcance se limitara al *homo sapiens*), o si tenemos una noción de “definición implícita” que no presuponga que sabemos qué *clase* de entidad es la que deseamos implícitamente definir. De ese modo, Lewis y yo elegimos (...) el camino de aceptar que existe un estado cerebral para cada actitud proposicional en el caso de cada organismo individual. Pero es aquí donde ambos fuimos víctimas de la “ciencia popular”. De manera que el funcionalismo no funciona. Esto es, no se ajusta a los fenómenos. Sin embargo, mucho es lo que aprendimos en el intento de ajustarnos a él»<sup>86</sup>.

### 1.3. La intencionalidad es ineliminable

Hemos visto hasta ahora que la intencionalidad es primitiva e irreductible, pero es que ninguna de estas dos características de la intencionalidad nos sirve para abordar la propuesta del eliminativismo, que resulta tentadora al proponer que la Psicología realmente Popular de los estados mentales es una teoría mítica, y que por lo tanto ha de ser reemplazada por una psicología científica. La tesis que Putnam propone como crítica a esta propuesta, y como defensa de la ineliminabilidad de la intencionalidad, afir-

ma que sólo podremos eliminar la intencionalidad cuando estemos dispuestos a realizar también esta eliminación con la Lógica Popular, y a eliminar el uso de una noción universal de verdad<sup>87</sup>.

Para demostrar esta tesis lo que hace Putnam es realizar una comparación entre la Psicología Popular y la Lógica Popular<sup>88</sup>. Lo que los Churchland y Stich vienen a proponer, es eliminar la primera por no ser científica, y en cambio aceptar las nociones de verdad y de referencia semántica de la segunda. Pero esta posición no es coherente, ya que «la idea de que la referencia y la verdad (o falsedad) son propiedades de las palabras y las oraciones en algo que merezca llamarse lenguaje, resultaría tan mitológica como la idea de que hay “actitudes proposicionales”»<sup>89</sup>.

Algunos piensan que este problema de la referencia y la verdad ha sido resuelto mediante la llamada teoría desentrecomilladora de la verdad. Esta se basa en la propuesta que hizo Tarski<sup>90</sup>, en su análisis de la verdad para los lenguajes formales, de enunciar la condición de verdad, para un lenguaje formal determinado, mediante oraciones de verdad del tipo: “la nieve es blanca” es verdadera si y sólo si la nieve es blanca. «Estas teorías de la verdad como desentrecomillado son importantes para la cuestión del eliminacionismo por la siguiente razón: si se pueden explicar la referencia y la verdad sin recurrir a la psicología, entonces la verdad y la referencia no constituyen un problema para las “ciencias cognitivas”. Si se las puede explicar en términos no psicológicos, entonces abandonar la psicología popular *no* implica abandonar las nociones de verdad y de referencia. El eliminacionismo no tiene que inmiscuirse en la lógica»<sup>91</sup>.

Esto es lo que Tarski pretendió, ya que los lenguajes formalizados se caracterizan por ser «lenguajes contruidos artificialmente en los que el sentido de cada expresión está determinado sin ambigüedad por su forma»<sup>92</sup>. Otra característica importante de estos lenguajes es distinguir el lenguaje objeto y el metalenguaje. El lenguaje objeto es un lenguaje formalizado del cual se conocen todos los signos que forman sus expresiones, sus oraciones cuentan con unas características estructurales que permiten distinguirlas del resto de las expresiones, es posible dar una lista de sus oraciones primitivas y contiene una serie de reglas de inferencia que permiten transformar unas oraciones en otras. El metalenguaje sí que contiene términos semánticos, pero que tendrán que ser definidos mediante términos no semánticos, además también debe contener todas las oraciones del lenguaje objeto (o una traducción de cada una de ellas) y un nombre para todas las expresiones del lenguaje objeto<sup>93</sup>.

Con estas premisas Tarski define el predicado ‘es una oración verdadera’ de forma recursiva para ese lenguaje formalizado. Esta recursión se

realiza considerando el tipo de conectivas con que se trabaja y comprobando si los predicados cumplen la relación de satisfacción. El concepto de satisfacción implica una relación entre predicados y  $n$ -tuplas de entidades para las cuales los predicados son verdaderos. La satisfacción es por tanto algo similar a lo que es la referencia para los predicados, pero con la diferencia de que la satisfacción esta limitada a una lista finita y fija de predicados.

La teoría tarskiana choca con uno de los grandes problemas que descubre toda teoría de la mente, que es el de dar razón de la dependencia del conocimiento humano de una realidad exterior al sujeto, es decir, de lo que se conoce como "el problema de la referencia". Todo el mundo advierte con facilidad que es el conocimiento sensible el que permite el acceso a otras realidades no sensibles, o el que podamos realizar juicios, etc. Pero este reconocimiento no impide que pueda considerar que ese conocimiento no tiene por qué ser contrastado con esa realidad externa para poder juzgar sobre su verdad, o para poder hablar de sinonimia entre dos términos. El intento de los eliminativistas es el de mostrar que es posible hablar de la verdad de una proposición sin hablar para nada de la realidad a la que esa proposición hace referencia. Pero esta es la gran paradoja de la teoría de la verdad tarskiana, ya que ese concepto de satisfacción y de verdad, al no ser conceptos generales, no sugieren cómo tratar un nuevo predicado que se incluya en la lista finita de predicados.

Estos problemas han llevado a Field<sup>94</sup> a afirmar que Tarski no define la verdad en términos no semánticos, sino que define las propiedades semánticas de las expresiones complejas en términos de las propiedades semánticas de las expresiones simples, ya que la teoría tarskiana es solamente una parte de una teoría completa que tendría además que contar con una teoría de la referencia para los nombres propios y los predicados. Por tanto si admitimos esta interpretación, la utilidad de la teoría desentrecomilladora queda claramente en entredicho.

Putnam considera dos tipos distintos de teorías desentrecomilladoras de la verdad. La primera teoría es la representada por Carnap, y que se caracteriza por definir la verdad en un lenguaje  $L$ , que consta de un número finito de oraciones, como la lista de oraciones de verdad, que satisfacen la construcción de Tarski, asociadas recursivamente con las oraciones del lenguaje  $L$ . Donald Davidson ha señalado que esta teoría hace que la verdad sea relativa al lenguaje, ya que «el interés de una teoría de la verdad (...) no consiste en decirnos lo que es la verdad en general, sino en que revela cómo la verdad de cada oración de un  $L$  particular depende de su estructura y constituyentes»<sup>95</sup>, es decir, reemplaza la noción "universal" e intuiti-



va de verdad por una serie infinita de nociones, “verdadera en L1,” “verdadera en L2”, pero «es sólo el hábito de *abreviar* el predicado de verdad con una expresión que contiene el nombre del lenguaje lo que *nos lleva a pensar que la relatividad de la noción de verdad respecto del lenguaje se halla reflejada de algún modo en la definición de verdad*»<sup>96</sup>. Esto es sólo una ilusión, ya que «una teoría de verdad para un lenguaje natural debe relativizar la verdad de una oración a las circunstancias de la emisión»<sup>97</sup>, y al hacerlo las condiciones de verdad dadas por la teoría tarskiana para esa oración, no tienen porqué traducir la oración descrita. Por tanto «la satisfacción de la Convención T de Tarski permanece como un *desiderátum* de una teoría, pero ya no se halla disponible como test formal»<sup>98</sup>. Esta teoría desentrecomilladora depende, por tanto, de cómo sean las cosas y de cómo es deletreada la oración, pero no de cuál sea el significado de la oración; es decir, todos los lenguajes que satisfagan la restricción sintáctica de un determinado lenguaje, también satisfacen la definición de verdad y referencia dada para ese lenguaje. La pretensión de eliminar el uso de una noción universal de verdad nos ha llevado a quedarnos también sin reglas semánticas, ya que la definición que hemos conseguido realizar de un lenguaje ha quedado reducida a sus reglas sintácticas<sup>99</sup>. Al definir la verdad como la lista de las oraciones de verdad que satisfacen la construcción de Tarski, hemos reducido la verdad en ese lenguaje a la adecuación de esa proposición con la lista de oraciones, por tanto lo importante no es lo que la proposición signifique sino la forma en que está escrita. Claramente no es este el concepto de verdad que quería lograrse, es por tanto necesario buscar otra teoría que defina la verdad de una forma diferente si no queremos acabar acudiendo de nuevo a las nociones de verdad y referencia que se querían reemplazar.

La segunda teoría desentrecomilladora de la verdad es llamada “desentrecomillado por desaparición”, y se caracteriza por considerar el desentrecomillado como sustituto de la noción de verdad. Es una teoría que rechaza «la idea de que las afirmaciones sean verdaderas porque corresponden a (“reflejan”, “representan adecuadamente”) la realidad (...) para todos aquellos casos en que no sea filosóficamente discutible (...). »<sup>100</sup>. Estos filósofos, (p. e., P. M. Churchland y R. Rorty), afirman que «*decir que una oración S es verdadera no es atribuir una propiedad más a S, sino simplemente “afirmar” S*»<sup>101</sup>. No es que pretendan negar la existencia de una propiedad normativa que las oraciones puedan o no tener, lo que quieren negar es que la noción clásica de verdad sea esa propiedad. Así para Rorty la verdad es algo que nosotros ponemos en las frases con las que estamos de acuerdo, es decir, que cumplen dos propiedades: «(1) que sean “correctas” según las pautas de nuestros iguales culturales (...) (2) que nos permitan “proceder con éxito”»<sup>102</sup>.



Rorty rechaza la referencia porque para él las personas pueden hacer referencia a cosas que no existen en la realidad, como pueden ser los personajes de ficción. Esto hace que el concepto de referencia sea para Rorty una construcción artificial, pero entonces «alguien que cree que Hamlet no se casa en la obra teatral, puede pensar que se refiere a un personaje de la obra teatral usando las palabras 'la esposa de Hamlet', pero esa persona se equivocaría al hacerlo (...). De acuerdo con Rorty, hablar realmente de algo implicaría su existencia»<sup>103</sup>, pero en el ejemplo que se ha puesto se ve que esto no tiene por qué ser cierto. No es fácil proponer una teoría de la referencia, pero lo que no puede hacerse es eliminarla si antes no se ha mostrado que no existe dicotomía entre los sucesos y la ficción.

En cambio, el "realismo científico" propondría que esta teoría desentrecomilladora de la verdad no necesita abandonar la idea de que las palabras y oraciones están relacionadas con cosas y sucesos en el mundo. Esto es consecuencia de que consideran la expresión de una oración como un signo de que van a suceder ciertas cosas, al igual que hacemos con un signo natural, pero no como una expresión que posea la propiedad de ser verdadera.

El problema de eliminar la idea de que la verdad sea una propiedad que posee la proposición, para pasar a considerar que la verdad es algo que se pone en ella o que es un signo de que van a suceder cosas, tiene el peligro de eliminar también el realismo<sup>104</sup>. Pero si en cambio lo que se pretende es mantener el realismo, esto es, la teoría de la verdad como correspondencia, compatibilizándolo con esta teoría desentrecomilladora, entonces es fácil acusar a ese realismo de no eliminar finalmente la intencionalidad, y por tanto de mantener una postura incoherente, ya que «rechazar las actitudes proposicionales por el hecho de que no pueden ser reducidas a propiedades físico-computacionales y, al mismo tiempo, conservar la relación de referencia (que, en este aspecto, se halla en la misma situación que las actitudes proposicionales) es inconsistente»<sup>105</sup>.

## 2. CRÍTICAS DE JOHN R. SEARLE

John R. Searle ha señalado también diversos problemas que conlleva el defender teorías reduccionistas. Algunos de estos problemas tienen que ver con el representacionismo, en especial con la carencia de semántica que caracteriza a toda teoría representacionista. En cambio, otros problemas tendrán que ver más con el tipo de estrategia empleada para estudiar el problema de la mente, ya que la elección de una determinada estrategia

puede comportar el perder fenómenos importantes que es necesario tener en consideración.

## 2.1. Los reduccionismos de la mente se basan en una ontología de la tercera persona

Un aspecto que distingue a muchas de las teorías reduccionistas es el intento de buscar un criterio observable de objetividad al abordar el problema que quieren estudiar. Esta estrategia, como se ve en el capítulo dos de la tesis, es muy útil para muchas ciencias, ya que permite cuantificar, e incluso predecir, los fenómenos que esa ciencia estudia. Searle señala que en el caso de la mente esta actitud entraña un grave error, ya que «una vez que asumimos que todo lo que es objetivo ha de ser accesible a cualquier observador, las preguntas dejan de hacer referencia a los estados mentales subjetivos y pasan a tratar de la objetividad del comportamiento externo. Y esto tiene la consecuencia de que en vez de preguntar “¿Qué es tener una creencia?”, “¿Qué es tener un deseo?”, “¿Qué es encontrarse en determinada clase de estado consciente?”, hacemos la pregunta de la tercera persona, “¿Bajo qué condiciones podríamos *atribuir*, desde fuera, creencias, deseos, etc., a *otro* sistema?”. Esto nos resulta perfectamente natural, ya que, por supuesto, muchas de las preguntas a las que debemos dar respuesta, acerca de los fenómenos mentales conciernen a otras personas y no sólo a nosotros mismos»<sup>106</sup>.

Este argumento de Searle parece recordar, en un principio, el acceso privilegiado a la propia mente que caracteriza a la filosofía cartesiana. Descartes lo que proponía era que los contenidos de nuestra mente están presentes a uno mismo siempre y de modo infalible. Aunque es cierto que esta doctrina, tal como Descartes la propone, no es defendida por ninguno de los filósofos importantes de lo mental, en cambio sí que se sigue notando su influencia, ya que con frecuencia se acepta que el acceso al contenido de la propia mente se realiza de forma inmediata y directa, e incluso de modo infalible. Esto ocurre con frecuencia al tratar los sentimientos, ya que muchos psicólogos consideran que el sentimiento es subjetivo, interno y que expresa de modo infalible nuestro modo de vivir las cosas. Así considerado, el sentimiento se opone a la sensación que sería el fenómeno objetivo, externo y que es falible por su valor cognoscitivo. Puede encontrarse este esquema conceptual en muchos filósofos de lo mental, y aunque no lo propongan para defenderlo, sí que lo aceptan como punto de partida de su estudio, y con ello consiguen, sin quererlo, que ese estudio que luego realizan, esté marcado con una premisa dualista que permanece implícitamente aunque se esté en contra del dualismo. Esta premisa dualista hace que el

objetivo que pretende lograrse, mediante ese estudio, sea demostrar que lo mental es tan objetivo como lo físico o que lo mental no existe.

Este acceso privilegiado puede comprenderse mejor si se emplea la metáfora de la visión. Lo que estaría proponiendo Descartes es la existencia de un sentido interno, cuyo objeto propio serían los contenidos de la mente, que permite ver lo interno con tal claridad que es imposible el error. Este sentido interno observaría los contenidos mentales de modo semejante a como se observan los objetos y sucesos con los sentidos externos, pero con la diferencia de que mientras para los sentidos externos pueden producirse errores en esa percepción, ya que el objeto puede estar lejos o mal iluminado o no ser familiar, etc., en cambio para el sentido interno estos problemas no pueden darse.

Cuando Searle habla de subjetividad, no se está refiriendo a este acceso privilegiado cartesiano al que llama "introspección", ya que lo que Searle quiere expresar al emplear el término subjetividad es el que «todas mis formas conscientes de intencionalidad, que me dan información acerca del mundo que es independiente de mí, están realizadas siempre desde un punto de vista especial»<sup>107</sup>. Y es ese punto de vista especial desde el cual yo recibo información acerca del mundo el que lleva consigo el que «el hecho de que tú y yo estemos teniendo una experiencia visual compartida del objeto no requiera que lo veamos bajo el mismo aspecto»<sup>108</sup>, y no es que no lo requiera, sino que es imposible que lo veamos bajo el mismo aspecto porque existe algo que no podemos compartir, que es el que ese aspecto o punto de vista con el que accedo a la realidad es mío, y esa subjetividad yo no puedo compartirla porque no tengo modo de acceder a ella, de observarla.

«El hecho fundamental de la subjetividad que estamos intentando observar hace que tal observación sea imposible. ¿Por qué? Porque en lo que se refiere a la subjetividad consciente, no hay distinción entre la observación y la cosa observada, entre la percepción y el objeto percibido. (...) Toda introspección que tengo de mi propio estado consciente es ella misma un estado consciente. (...) Es afirmar simplemente que el modelo estándar de observación no funciona para la conciencia subjetiva. No funciona tanto para la conciencia de otros como para la de uno mismo. Por este motivo, la idea de que podía existir un método especial para investigar la conciencia, llamado "introspección", que se supone que es una clase de observación interna, estaba condenada al fracaso desde el inicio»<sup>109</sup>. No es posible, por tanto, un conocimiento que sea sólo objetivo (de la tercera persona) de mi propia conciencia, a no ser que esté dispuesto a perder el carácter subjetivo que la conciencia posee.



A esta teoría que Searle critica se le conoce como “fisicalismo” o “naturalismo” y se caracteriza por «demandar que el discurso mentalístico sea reducido, explicado y eliminado en favor o de un discurso que sea “aceptable”, o mediante visiones ya descubiertas en las ciencias físicas o naturales»<sup>110</sup>. Por tanto, si se afirma, como hacen las teorías reduccionistas, que los fenómenos mentales pueden ser reducidos a fenómenos observables, la consecuencia inmediata de este postulado, será que la epistemología empleada será una epistemología en tercera persona, y por tanto se rechazará cualquier intento (por innecesario) de hacer una epistemología en primera persona. Esto comporta el que «el modo en que se aplica, en la práctica, el punto de vista de la tercera persona, hace difícil que veamos la diferencia entre algo que tiene realmente mente, como por ejemplo un ser humano, y algo que se comporta *como si* tuviera una mente, como es un ordenador»<sup>111</sup>.

Parece, por tanto, que lo que propone Searle es una *asimetría epistémica* entre la forma de conocer mis propios estados y los estados de los otros. «Si intento observar la conciencia de otro, lo que observamos no es su subjetividad sino solamente su comportamiento consciente, su estructura y las relaciones causales entre la estructura y el comportamiento (...). Es decir, no hay forma en que yo pueda observar la conciencia de otro como tal conciencia, sino que más bien le observo a él, a su comportamiento, y a las relaciones entre él, la estructura, el comportamiento y el entorno»<sup>112</sup>. Se podría pensar que Searle está aceptando con esto un tipo de acceso privilegiado del tipo de la metáfora de la habitación, que sugiere que mi conciencia es como una habitación privada a la que sólo yo tengo acceso, es decir, que la conciencia es un lugar al que sólo yo puedo acceder. Pero ya ha señalado Searle que yo no puedo observar mi subjetividad, es decir, yo no tengo un acceso privilegiado a mi subjetividad, no tanto porque mi forma de acercarme a ella sea idéntica a la de un tercero, sino porque yo no puedo observar mi subjetividad como tampoco lo puede hacer un tercero. Esta metáfora supone la existencia de una realidad en la que ‘yo’ entro que es distinta del ‘yo’ que entra en ella, divide por lo tanto al ser humano al menos en dos, y pierde de esta forma una de las características más importantes del ser humano, que es la de su unidad. Searle no cae en esta trampa y propone un tipo de subjetividad que no se ajusta a ninguna de esas dos metáforas, pero que encuentra un gran rechazo en aquellos filósofos que quieren lograr un estudio objetivo de la realidad de la mente.

¿Por qué hay filósofos que niegan la existencia de estados intrínsecamente subjetivos? Searle considera que hay cuatro rasgos de los fenómenos mentales que han hecho que sea difícil encajar los fenómenos mentales



en la concepción ‘científica’ que se tiene del mundo<sup>113</sup>. El primero de ellos sería la conciencia, ya que es difícil ver cómo sistemas meramente físicos pueden tener conciencia. El segundo sería la ‘intencionalidad’, es decir, el rasgo según el cual mis estados mentales se dirigen, o refieren, a objetos distintos de mí mismo, entonces surge la pregunta: ¿Cómo puede ser que la materia que está dentro de mi cerebro se refiera a algo? El tercero es el de la subjetividad de los estados mentales, es decir, que yo puedo sentir mis dolores y tú no puedes, o lo que es lo mismo, que yo soy consciente de mí mismo y de mis estados mentales, como algo completamente distinto de los míos y de los estados mentales de las otras personas. Por último está el problema de la causación mental que podría resumirse en la pregunta: ¿Cómo puede algo mental tener una influencia física?, si se acepta que la mente no tiene conexión alguna con el cerebro, o que la mente es un epifenómeno del cerebro, entonces la mente es tan poco importante causalmente como lo es la espuma de la ola para causar el movimiento de la ola.

Searle propone, como ejemplo de este tipo de actitud objetiva que rechaza la existencia de estados intrínsecamente subjetivos, el caso del cambio que se realiza en la pregunta acerca de la conciencia. «La pregunta es: ¿Cómo es que trozos de materia no inteligentes producen inteligencia? Lo primero que deberíamos captar es la forma de la pregunta. Por qué no hacemos la pregunta de un modo más tradicional: ¿Cómo es que trozos de materia no conscientes producen conciencia? (...) Pero en esta época no somos partidarios de hacer la pregunta de este modo ya que necesitamos un criterio “objetivo” de conciencia (...). Pensamos que para la inteligencia tenemos un criterio objetivo e impersonal. Pero inmediatamente encontramos una dificultad (...) ya que si la inteligencia es definida de forma conductista, entonces no es fácil afirmar que las neuronas no sean inteligentes. Las neuronas, al igual que las otras cosas que hay en el mundo, se comportan según pautas regulares y predecibles. (...) Si la inteligencia es definida conductistamente, entonces las neuronas son realmente inteligentes (...). Entonces la pregunta contiene una premisa que en sus mismos términos es falsa. La pregunta presupone falsamente que esos elementos no cumplen el criterio de tener inteligencia»<sup>114</sup>.

Searle concede una especial relevancia a la pregunta acerca de la conciencia debido a que considera que existe una conexión estrecha entre la conciencia y la intencionalidad: «Sólo un ser que pueda tener estados intencionales conscientes, puede tener los otros estados intencionales»<sup>115</sup>. «Esta tesis tiene enormes consecuencias para el estudio de la mente. Implica, por ejemplo, que toda discusión de la intencionalidad que deja fuera la cuestión acerca de la conciencia será incompleta. Es posible describir la estructura

lógica del fenómeno intencional sin discutir la conciencia, (...) pero existe una conexión conceptual entre la conciencia y la intencionalidad cuya consecuencia es que una teoría completa de la intencionalidad requiere tener presente la conciencia»<sup>116</sup>.

El gran problema que implica el abandono de la subjetividad, a la hora de dar una explicación de los fenómenos mentales, es que «una vez que hemos adoptado esta estrategia (que ha sido la más usada en los últimos cincuenta años de la filosofía de la mente), entonces se hace imposible describir la conciencia, ya que se hace literalmente imposible reconocer la subjetividad de la conciencia. (...) Estas propuestas presuponen una concepción de la realidad del tipo de la tercera persona, una concepción de la realidad que no es sólo epistemológicamente objetiva sino más bien ontológicamente objetiva; y tal realidad no posee un lugar para la conciencia ya que no posee un sitio para la subjetividad ontológica»<sup>117</sup>. Esto ya lo habían afirmado anteriormente otros filósofos: «Fundamentalmente un organismo tiene estados mentales conscientes sí y sólo sí existe algo que es *ser* ese organismo (...). Podemos llamar a ese algo el carácter subjetivo de la experiencia. Este no es capturado por ninguno de los recientes y familiares análisis reductivos de lo mental, ya que todos ellos son lógicamente compatibles con su ausencia (...). Todo programa reduccionista tiene que basarse en el análisis de lo que debe ser reducido. Si ese análisis deja algo fuera, el problema ha sido planteado falsamente»<sup>118</sup>.

Searle, al igual que Putnam, reacciona en contra de este intento de reducción, y desarrolla un argumento empleado anteriormente por otros filósofos<sup>119</sup>. «Considera cuáles son los hechos que en el mundo hacen que tú estés ahora en un cierto estado consciente como es el dolor (...). El primero y más importante, es el hecho de que estás teniendo unas desagradables sensaciones conscientes, y que las estás experimentando desde tu subjetividad, desde el punto de vista de la primera persona. (...) Pero el dolor también es causado por unos procesos neurológicos que consisten en gran parte en pautas de disparos de neuronas en el tálamo y otras regiones de tu cerebro. (...) Supón que intentamos decir que el dolor realmente “no es otra cosa más que” las pautas de disparos de las neuronas. Bien, si nosotros intentamos tal reducción ontológica, las características esenciales del dolor son dejadas de lado. Ninguna descripción objetiva, del tipo de la tercera persona, de los hechos fisiológicos abarcaría la subjetividad, el carácter de primera persona del dolor, ya que simplemente las características de la primera persona son diferentes de los de la tercera persona. (...) Este hecho tiene obvias consecuencias epistémicas: el conocimiento de que yo tengo dolor tiene un tipo de base diferente respecto a mi conocimiento de que tú

tienes dolor. Pero el punto antirreduccionista del argumento es ontológico y no epistémico»<sup>120</sup>.

Con esto lo que quiere Searle afirmar es que no es posible lograr una reducción de los fenómenos mentales similar a la que realiza la física. En la física se inicia el estudio del fenómeno por medio de las características observables e incluso partiendo de apariencias subjetivas (p. e., “rojo” es lo que parece rojo a observadores normales bajo condiciones normales), posteriormente se redefine el fenómeno eliminando toda referencia a la subjetividad, ya que lo que nos interesa es explicarlo mediante sus causas físicas. No es que con el reduccionismo físico «primero descubramos todos los hechos y luego descubramos uno nuevo, es decir, el hecho de que sea reducible; sino que simplemente redefinimos el calor de forma que la reducción se obtiene de la definición. Pero esta redefinición no elimina, ni estaba pensada para eliminar, las experiencias subjetivas del calor (...). Estas siguen existiendo como siempre»<sup>121</sup>. Estas reducciones no demuestran que esos fenómenos físicos no existan realmente tal como son percibidos, es decir, mediante sus características superficiales, sino que mediante su redefinición por medio de las causas que subyacen al fenómeno se eliminan las experiencias subjetivas de esta definición. Pero esto no puede aplicarse al fenómeno de la conciencia ya que lo que interesa del mismo no son sus causas físicas sino la experiencia subjetiva que produce. No es que no se pueda hacer esa redefinición del fenómeno mental a sus causas físicas, «pero por supuesto, la reducción del dolor a su realidad física todavía deja sin reducir la experiencia subjetiva del dolor»<sup>122</sup>. Lo que quiere afirmar con todo esto es que en el caso de la física «es una característica general de tal reducción que el fenómeno sea definido en términos de la “realidad” y no de las “apariencias”. Nosotros no podemos hacer esta clase de distinción entre apariencia y realidad en el caso de la conciencia ya que la conciencia consiste ella misma en las apariencias. *Donde la apariencia esta involucrada no podemos hacer la distinción entre apariencia y realidad, ya que la apariencia es la realidad*»<sup>123</sup>.

Searle resume este argumento de la siguiente forma: «El contraste entre la reductibilidad del calor, el color, la solidez, etc., por un lado, y la irreductibilidad de los estados de conciencia, por el otro, no reflejan ninguna distinción en la estructura de la realidad, sino una distinción en nuestra práctica definicional. (...) Nosotros no eliminamos realmente la subjetividad del rojo, por ejemplo, cuando reducimos el rojo a reflectancias de luz; lo que hacemos es dejar de hablar de la parte subjetiva del “rojo”. No eliminamos ningún fenómeno subjetivo cualquiera con estas “reducciones”; simplemente dejamos de hablar de ellos. (...) Nos encontramos en un uni-



verso que posee, como componente de la realidad física, una componente física subjetiva irreducible»<sup>124</sup>.

Por un lado, tenemos que Searle está afirmando que lo mental es físico, ya que «cuando yo digo que la conciencia es una característica física de alto nivel (...) lo que realmente quiero decir es que la conciencia *qua* conciencia, *qua* mental, *qua* subjetiva, *qua* cualitativa es física, y física *porque es mental*»<sup>125</sup>, pero por el otro considera que lo mental es irreductiblemente subjetivo. Searle no entiende por 'físico' la *res extensa* cartesiana ya que, de ser así, su teoría de que todo es físico sería fácilmente refutada por la física moderna<sup>126</sup>. Más bien parece que por físico entiende las partículas físicas, es decir, las diversas clases de entidades que reconoce la física moderna, pero este empleo de conceptos físicos no implica una fe ciega en la física, ya que reconoce que no se puede suponer que la física moderna sea totalmente correcta<sup>127</sup>, y que por lo tanto estas partículas podrían ser otras si la física demuestra que son otras. Pero lo que interesa destacar es que Searle parece afirmar que lo mental está constituido por partículas físicas, al mismo tiempo que defiende que las propiedades mentales no pueden ser reducidas a propiedades físicas.

Si la mente está constituida por partículas físicas, ¿por qué entonces sus propiedades no pueden ser explicadas por la física? Searle va a responder a esta objeción afirmando que el reduccionismo causal no implica el reduccionismo ontológico, es decir, que lo que interesa no son las causas físicas que provocan un determinado fenómeno, sino el que ese fenómeno sea subjetivo. Searle lo que quiere afirmar es que si consideramos la causalidad, yo puedo explicar perfectamente la causalidad mental mediante la causalidad física, pero que eso no me permite realizar una reducción ontológica de la subjetividad del fenómeno a esa materia que la causa. Rechaza, por tanto, la explicación neurológica por no dar cuenta de esa subjetividad.

Pero ¿qué es exactamente esa subjetividad que no es posible reducir? Searle parece aceptar la analogía de las propiedades 'superficiales' de la física, como es por ejemplo la de la liquidez del agua: «[la conciencia es una] característica o propiedad del cerebro en el sentido, por ejemplo, en el que la liquidez es una característica del agua»<sup>128</sup>. La liquidez del agua está constituida por la microestructura del agua, por lo tanto no hay nada nuevo en lo que respecta a las partículas que la constituyen, pero tampoco es idéntica la liquidez con la forma en que esas moléculas se estructuran. Searle acepta las tesis sobre la superveniencia de lo mental sobre lo físico, pero entonces ¿cómo compaginar esta superveniencia con la eficacia causal de lo mental? Para Searle no existe ningún problema: «El fenómeno de que las características mentales sean supervenientes sobre las características neuro-



nales, en ningún momento disminuye su eficacia causal. La solidez del pistón es causalmente superveniente sobre su estructura molecular, pero esto no hace que la solidez sea epifenoménica; de modo similar, la superveniencia causal de mi actual dolor de espalda en los microsucesos que ocurren en mi cerebro, no hacen que mi dolor sea epifenoménico»<sup>129</sup>.

A primera vista, lo que parece es que si yo puedo explicar mi actual dolor de espalda mediante los poderes causales de mis estados neuronales, es difícil que no considere que mi dolor es epifenoménico. Pero Searle acepta que yo puedo explicar causalmente el proceso mental ya que «nosotros podríamos simplemente definir, por ejemplo, el dolor como “patrones” de actividad neuronal que causan las sensaciones subjetivas de dolor (...) pero, por supuesto, la reducción del dolor a su realidad física todavía deja la experiencia subjetiva del dolor sin reducir»<sup>130</sup>, es esa realidad subjetiva la que el reduccionismo causal no puede reducir. Este componente indexical que añadimos al fenómeno mental es el que hace que, aunque la causalidad del fenómeno mental pueda ser explicada mediante la causalidad del fenómeno neuronal, existe al menos un rasgo que la descripción neuronal no puede explicar, y es que ese fenómeno es mío. Fenómeno que sólo puedo captar si acepto como método de estudio una ontología de la primera persona.

Este punto de vista de la primera persona resulta importante a la hora de intentar comprender el experimento de la “habitación china”<sup>131</sup> que Searle ha hecho famoso en los últimos años. Si se supone que en una habitación se pone a una persona y se le suministran caracteres ideográficos chinos y unas reglas para que, dada una cadena de caracteres ideográficos como entrada, él pueda, siguiendo las instrucciones que le proporcionan las reglas, dar unos caracteres a la salida, entonces puede ocurrir que para la persona que esta fuera le parezca que la persona que está en la habitación entiende el chino.

Desde el punto de vista del observador externo que se halla fuera de la habitación, parece obvio que la persona que está dentro entiende el chino. Pero en cambio desde el punto de vista de la primera persona, la persona que está dentro de la habitación se da cuenta de que no entiende la oración que genera como salida. ¿Cuál de los dos está en lo cierto? Es fácil que uno responda inmediatamente que la persona no entiende el chino, pero es necesario argumentar un poco más la respuesta. El primer problema que surge, si se da esta respuesta, es el de explicar ese acceso privilegiado que tiene el agente a sus propios estados, ya que parece que yo tengo un conocimiento del cual carece la persona que me observa. Searle ha mantenido que él está en contra de este acceso privilegiado, por lo tanto es necesario explicar esta

asimetría de otro modo. En el caso en estudio, la asimetría proviene más bien de que el observador carece de los datos que la persona que está dentro de la habitación posee acerca de lo que está haciendo. El observador externo lo que ve es sólo una parte de su comportamiento, aquel que el observador es incapaz de distinguir de la de otra persona que sí entienda el chino, por tanto desde el punto de vista desde el cual él lo estudia no tiene forma de saber si esa persona entiende el significado o no de esa oración, ya que se comporta como si realmente lo entendiera. En cambio, la persona que está dentro se da perfecta cuenta de que lo que está haciendo es tomar los caracteres de entradas y aplicar unas leyes, sin importarle lo más mínimo lo que significa la oración que da como salida<sup>132</sup>. Tanto es así que si el observador le introduce una oración sin sentido, es casi seguro, a no ser que el conjunto de reglas que le hayan dado consiga detectar esa irregularidad, que le responda con otra oración sin sentido.

Los defensores de las teorías reduccionistas ven objeciones a este experimento. Y así Dennett<sup>133</sup> aunque reconoce que tal como Searle plantea el experimento, no hay ninguna dificultad en aceptar que «no podría decirse que nadie ni ninguna de las cosas entiende chino»<sup>134</sup>, al mismo tiempo propone modificar un poco el experimento y considerar que «nuestro héroe en la habitación china “internaliza todos esos elementos del sistema”, es decir, “incorpora todo el sistema”»<sup>135</sup>, entonces ese nuevo ser, para Dennett, entiende chino. Pero con toda esta argumentación Dennett lo que hace es volver a situar el problema en la posición que Searle ha intentado criticar desde un principio. Está claro que si lo que yo considero es el sistema total desde el punto de vista de una tercera persona, con facilidad acabaré concluyendo que la habitación entiende chino, pero es que el error está en considerar que ése es el único punto de vista con el cual se ha de estudiar el problema. Pero si la situación sigue siendo la misma, ¿qué es lo que diferencia a la persona que ha internalizado ese conjunto de reglas, de la persona que las tiene escritas sobre un papel? Para Dennett la diferencia es total, la primera entiende chino y la segunda no. Para Searle la diferencia sería mínima, ya que una de las personas ha logrado aprenderse las reglas y la otra no, pero ese aprendizaje de las reglas no implica que la persona entienda el chino, ya que entender es algo más que manipular unas reglas, es querer significar algo, pretender hacer referencia a una realidad que está más allá del propio lenguaje y de las reglas con que se construyen las oraciones de ese lenguaje. «No hay ninguna manera de que el sistema pueda obtener a partir de la sintaxis la semántica»<sup>136</sup>. El problema de Dennett está en que critica la posición de Searle desde su visión *objetiva* (científica), pero es que Searle ya le ha reconocido que desde ella no se puede entender su experimento. Además, pienso que es conveniente señalar que Dennett

dota al ordenador de unas capacidades que él supone que posee desde el punto de vista de la tercera persona, pero que desde el punto de vista de la primera persona pueden ponerse muy en duda. Por tanto es necesario mostrar que las tiene, pero como esa demostración exige adoptar un punto de vista de la primera persona, pienso que Dennett nunca lo demostrará.

Además, lo que los reduccionistas y eliminativistas pretenden no es sólo no aceptar el que el estudio de este experimento pueda realizarse de alguna forma desde el punto de vista de la primera persona, sino también defender que el caso del ordenador tiene poco que ver con el experimento de la habitación china de Searle<sup>137</sup>, ya que en el caso de la habitación china el hombre que está dentro de la habitación está considerando que se aburre mucho con semejante experimento, y esto hace que se deba considerar que no es capaz de implementar el programa que ejecuta porque no lo hace algo propio, sino que lo considera algo exterior a él; en el caso del ordenador esto no pasa. Un ejemplo de esta posición es la afirmación que hace Hayes de que «los ordenadores actualmente no *leen*, ni *obedecen*, ni *siguen* reglas. Más bien, cuando una ‘regla’ es trasladada a sus procesadores, estos se transforman en un estado en el cual los cambios que *causan* pueden ser apropiadamente interpretados como la consecuencia de la regla. Pero, tal como todos nosotros sabemos, no hay ningún pequeño hombre dentro del ordenador que esté leyendo y siguiendo algo, y toda esta forma de hablar es solamente una metáfora que cualquiera, con un pequeño conocimiento de la sofisticación computacional, aprende a manejar»<sup>138</sup>. He de reconocer que mi experiencia en el diseño de ordenadores me lleva a disentir fuertemente de esta posición que Hayes defiende<sup>139</sup>.

El problema que sí que plantea esta teoría de Searle es la de dar respuesta a preguntas del tipo: ¿cómo puedo conocer el significado de lo que dice otra persona? o ¿cómo sé que lo que dice otra persona tiene significado? Searle reconoce que existe una asimetría entre lo que yo conozco de mí mismo y lo que puedo conocer de los otros, pero al mismo tiempo afirma que «los sucesos lingüísticamente relevantes deben ser los mismos en las preguntas: ¿qué entiendo cuando otra persona está diciendo “Llueve”? y ¿qué es para mí entenderme cuando estoy diciendo “Llueve”?», ya que (...) lo que yo consigo cuando le entiendo es exactamente lo que él consigue cuando me entiende. Pero si yo ya me entiendo, todo lo que podré aprender de estudiar su caso lo puedo aprender de estudiar el mío»<sup>140</sup>. Pero parece que esto implica un cierto solipsismo, no tanto un aislarme de la realidad externa, sino un acceso a la realidad que sólo en mí mismo puedo realizar, porque sólo en mí mismo podré descubrirlo. «El suceso crucial en cuestión es que la realización de actos del habla -y el significar cosas con expresio-



nes- se produce a un nivel de intencionalidad intrínseca de primera persona»<sup>141</sup>. «Imagina que el hablante emite la oración “está lloviendo”. El hacer esta emisión es una acción intencional (...) pero así caracterizada, esta intención no lleva consigo todavía ningún significado. ¿Qué es lo que se añade cuando decimos que el hablante no sólo *emitió* intencionalmente la oración “está lloviendo”, sino que quiso *significar* que estaba lloviendo?»<sup>142</sup>. «La diferencia entre decir que está lloviendo y significarlo, y decirlo pero sin significarlo se puede obtener examinando la pregunta, ¿Qué se considera como un error?»<sup>143</sup>. Con esta afirmación lo que Searle quiere decir es que cuando una persona busca significar P diciendo x, lo que pretende es que, decir x sea entendido como el estado de las cosas de P<sup>144</sup>, es decir, que cuando esa persona está diciendo x lo que busca conseguir es que la persona que le escuche piense P. Por lo tanto, la forma de conocer lo que es el significado es lograr determinar cuándo una persona consigue adecuar su emisión al estado de las cosas, y por eso Searle propone que la diferencia entre significar y no significar se puede descubrir estudiando el error, es decir, descubriendo cuándo no se significa.

Inmediatamente surge la contrarréplica de afirmar que con frecuencia las personas se equivocan y en cambio su oración posee significado. Pienso que es bueno entender lo que Searle pretende decir al hacer estas afirmaciones. ¿Qué es lo que diferencia a una persona que dice una oración con significado y de otra persona que dice esa misma oración pero sin significado? Pues que la primera busca que la oración que dice se corresponda con un estado de las cosas que es exterior a la oración, mientras que la otra no pretende tal cosa. No afirma Searle que una persona logre significar porque consigue que su oración represente sin error ese estado de las cosas, pero sí que la búsqueda de esas condiciones de satisfacción entre lo que digo y el estado de las cosas hace que mi oración signifique algo. Pero entonces, ¿cómo sé yo que una persona realmente está diciendo una frase con significado?

La respuesta de Searle es que no puedo saberlo, sólo puedo suponerlo, ya que el problema del significado, tal como Searle lo estudia, tiene que ser visto desde una perspectiva de la primera persona. Eso quiere decir que cuando Searle habla de significar o no significar, él está pensando en su propio caso. «El lenguaje es verdaderamente público; y no es una cuestión de significados como entidades de introspección, de objetos privados, o de acceso privilegiado, u otra cualquiera de la parafernalia cartesiana. El punto importante, sin embargo, es que cuando entendemos a alguien o a nosotros mismos, lo que necesitamos -entre otras cosas- es un conocimiento de contenidos intencionales. El conocimiento de estos contenidos no es equiva-



lente al conocimiento que nos permite el emparejamiento del comportamiento público con los estímulos, ni tampoco del emparejamiento de lo que se dice con las condiciones del mundo. Esto se descubre de forma especial en el caso de la primera persona, y nuestro rechazo del caso de la primera persona nos conduce a una mala comprensión del lenguaje»<sup>145</sup>. Searle, por lo tanto, reconoce que existe una indeterminación en la interpretación semántica de las otras mentes, pero es que esa indeterminación es consecuencia de que yo no puedo aplicar el punto de vista de la primera persona a esas otras mentes. El único camino para saber si una oración se dice con significado exige emplear un punto de vista de la primera persona, ya que de no hacerlo así se dejan de lado los contenidos intencionales, y es este el error que han cometido todos aquellos que han intentado objetivizar el estudio del significado y que les ha llevado a una mala comprensión del lenguaje. Yo puedo afirmar que lo que dicen los otros tiene significado porque supongo que la emisión de la oración que dicen cuenta con los contenidos intencionales con los que yo dotaría a esa misma oración si la dijera, pero no porque posea un test o una teoría que me permita comprobar cuándo una oración está dotada de significado y cuándo no lo está, porque entre otras cosas, tal test o teoría no existe, a no ser que se quiera dejar de lado el aspecto subjetivo de los contenidos intencionales. Pienso que Searle ha logrado mostrar que tal eliminación conlleva un grave error.

## 2.2. La mente no es un ordenador digital

Históricamente, el considerar que la mente es un ordenador digital se remonta a las propuestas de Turing<sup>146</sup>, que sostenían que existe, para todo algoritmo, una máquina de Turing que puede implementarlo, y que existe un máquina universal de Turing que puede simular cualquiera de las máquinas de Turing. Pero «¿existen buenas razones para suponer que el cerebro deba ser una máquina universal de Turing?»<sup>147</sup>. La respuesta afirmativa a esta pregunta ha dado lugar a lo que se conoce como ciencia cognitiva, que intenta descubrir los programas que posee el cerebro a través de su implementación en programas de ordenador. La identidad entre ambos tipos de programas se logra cuando el ordenador reproduce la acción del ordenador humano y los psicólogos encuentran evidencias de que los procesos internos son los mismos en los dos tipos de ordenadores.

Pero se pregunta Searle: «¿Cuáles son exactamente las características anatómicas y fisiológicas de los cerebros que se van a tener en cuenta? ¿Qué es exactamente un ordenador digital? ¿Cómo se conectan las respuestas a estas dos preguntas? (...) Para mi sorpresa, al leer esta literatura he

encontrado que parece existir un peculiar hiato filosófico ante estas preguntas<sup>148</sup>. Esta denuncia que realiza Searle lo que busca señalar es que la ciencia cognitiva presenta por un lado teoremas, como son el de Turing y el de Church, y una teoría de la función recursiva que provienen de la matemática, y por el otro un conjunto de dispositivos electrónicos. Pero sorprendentemente nadie por ahora ha intentado realizar el trabajo filosófico de conectar la matemática con la electrónica. Lo que ocurre es que muchos de los que participan en el desarrollo de la ciencia cognitiva lo que buscan es lograr realizar que una máquina se comporte de forma inteligente. Esto hace que se preocupen, no tanto de buscar definiciones que nos permitan determinar lo que es un símbolo, o un algoritmo, o un ordenador digital, sino más bien de la caracterización de los procesos que son clasificados como inteligentes, abstrayéndolos de su implementación física<sup>149</sup>.

Esto hace que al intentar responder a la pregunta acerca de lo que es un ordenador digital, se descubre enseguida que la realización material, es decir, el *hardware* de tal ordenador, es irrelevante para la descripción computacional. Entonces, «¿por qué los defensores del computacionalismo no están preocupados con las implicaciones de esta múltiple realizabilidad (del sistema)? La respuesta es que ellos piensan que es típico de la tesis funcional el que la misma función admita múltiples realizaciones. Con esta idea, los ordenadores son semejantes a los carburadores y a los termostatos. Así como los carburadores pueden estar hechos de bronce o de acero, los ordenadores pueden estar hechos con un rango infinito de materiales»<sup>150</sup>.

Pero realmente sí que existe una diferencia, ya que «los carburadores y termostatos están definidos en términos de la producción de ciertos efectos físicos. (...) En cambio, los ordenadores están definidos sintácticamente en términos de la *asignación* de ceros y unos»<sup>151</sup>, y «la "sintaxis" no nombra una propiedad física, como lo hace la masa o la gravedad»<sup>152</sup>. Por tanto, Searle afirma que, en el caso de los ordenadores, el poder lograr que el proceso pueda ser implementado de forma múltiple se logra a través de sus propiedades sintácticas, y no porque se alcance el efecto físico que deseamos, (en el caso del termostato que abra o cierre un circuito eléctrico, o cierra una llave de paso, etc). Así expresada, esta teoría tiene una fácil refutación, por eso es necesario profundizar un poco más en ella y descubrir qué es lo que realmente está queriendo decir Searle.

«Esto tiene dos consecuencias que son desastrosas:

1. El mismo principio que implica la posibilidad de realización múltiple del sistema, parece que implicaría la posibilidad de su realización universal. Si la computación se define en términos de la asignación de sintaxis,

entonces cualquier cosa podría ser un ordenador, ya que a cualquier objeto se le pueden hacer adscripciones sintácticas.

2. Todavía peor es el hecho de que la sintaxis no es intrínseca a la física. La adscripción de propiedades sintácticas es siempre relativa a un agente u observador que trata ciertos fenómenos físicos como sintácticos»<sup>153</sup>.

La primera dificultad podría salvarse si se restringe la definición de computación haciéndola depender por ejemplo de las relaciones causales entre los estados de los programas, de la posibilidad de programación y control de los mecanismos, y de su situación en el mundo real. Pero estas «restricciones en la definición de computación no sirven de ayuda en la siguiente discusión, ya que el verdadero problema de fondo es el debido a que la sintaxis es esencialmente una noción relativa a un observador. La realización múltiple de procesos computacionales equivalentes en diferentes medios físicos no es sólo un signo de que los procesos son abstractos, sino de que estos procesos no son intrínsecos al sistema. Dependen de una interpretación desde el exterior»<sup>154</sup>. «Los estados computacionales no se descubren dentro de lo físico, sino que son asignados a lo físico»<sup>155</sup>.

Searle plantea, por lo tanto, un nuevo argumento distinto del que ya había propuesto con el argumento de la habitación china<sup>156</sup>. Hace diez años él había afirmado que, en el caso de los ordenadores, la semántica es extrínseca a la sintaxis, pero ahora va todavía más al fondo y afirma que, a su vez, la sintaxis, en los ordenadores, es extrínseca a lo físico, es decir, al material sobre el cual esta sintaxis se implementa. «Para entender bien este argumento es necesario comprender la distinción entre las características del mundo que son *intrínsecas* y aquellas que son *relativas al observador*»<sup>157</sup>. Esta distinción entre características intrínsecas y extrínsecas plantea diversos problemas epistemológicos que no voy a resolver ahora. Pienso que es suficiente considerar que las realidades externas a mí, que no dependen de mi forma de definir las, seguirán siendo lo que eran las defina como las defina, incluso aunque me equivocara al hacer esa definición, ya que no puedo actuar con mi conocimiento sobre algo exterior a mí mismo. Como ejemplos de esas realidades que son intrínsecas al mundo propone la “masa”, la “atracción de la gravedad” y la “molécula”, que son realidades que seguirían existiendo como tales aunque nunca hubiera existido alguien que las observara y diera una definición de ellas. En cambio, si yo a esas realidades les añado como algo propio la definición que yo hago de ellas, entonces las mismas realidades que hemos considerado antes o “un día estupendo para ir de excursión”, una “silla” o una “manguera”, existirían como tales porque existe un observador que es capaz de juzgar si poseen esa propiedad



que les hace ser tal cosa. Esto no significa que esta asignación sea arbitraria, sino que tal asignación está relacionada con aquel observador que usa u observa esa realidad como tal. Esto no quiere decir que esa definición no recoja propiedades intrínsecas de esa realidad, sino que puede poner propiedades en esa realidad que no son intrínsecas.

«La pretensión de la ciencia natural es la de descubrir y caracterizar las realidades que son intrínsecas al mundo natural. Por la propia definición de computación y cognición, no hay forma alguna de que la ciencia cognitiva computacional pueda ser una ciencia natural, ya que la computación no es una característica intrínseca al mundo. Es asignada de forma relativa por los observadores»<sup>158</sup>. Searle quiere afirmar que la computación es una propiedad que es añadida a esa realidad física, no es, por tanto, algo que esa realidad posea por sí misma, sino que esa propiedad la adquiere en el momento en el que se le añade mediante la definición que un observador hace de ella.

Esta respuesta lleva consigo consecuencias diversas. Si yo acepto que, para dar una explicación computacional del proceso físico realizado por un ordenador, necesito un observador que ponga esa característica computacional en ese proceso físico, entonces «falla el intento de eliminar la falacia del homúnculo»<sup>159</sup> mediante la descomposición recursiva (del proceso), ya que la única forma de hacer intrínseca la sintaxis a lo físico es poniendo un homúnculo en lo físico»<sup>160</sup>. Algunos autores, como Dennett<sup>161</sup>, han afirmado que el homúnculo puede desecharse. Su propuesta es analizar las operaciones del ordenador en elementos cada vez más simples, hasta alcanzar el simple *flip-flop*<sup>162</sup>, y considerar que con este análisis el homúnculo de alto nivel se convierte progresivamente en un homúnculo más estúpido, hasta alcanzar el simple *flip-flop* en el cual no hay realmente homúnculo. Es decir, los cognitivistas están dispuestos a aceptar que los niveles superiores de abstracción de un proceso computacional son relativos al observador que interpreta el proceso como una operación determinada; pero en cambio, a medida que recursivamente descendemos hacia niveles inferiores y alcanzamos una descripción basada en ceros y unos, negarán que tal descripción sea interpretativa, es decir, relativa al observador. «Pero conceder que los niveles superiores de computación no son intrínsecos a la física es también conceder que tampoco lo son los niveles inferiores»<sup>163</sup>.

La búsqueda de una teoría que permita desechar el homúnculo sigue siendo consecuencia del intento de defender un punto de vista de la tercera persona. Desde ese punto de vista los ceros y unos que caracterizan a los actuales ordenadores son algo intrínseco a la física, no los ceros y unos sino



los voltajes que se interpretan como tales números. Esos voltajes son algo que el ordenador posee por sí mismo, y en eso Searle está de acuerdo, lo que no puede aceptar es que se le intenta dar a esos voltajes unas características que no pueden poseer. Si se quiere que esos voltajes tengan capacidades sintácticas, no queda más remedio que considerar que existe un homúnculo al nivel del sustrato físico que hace que la razón de ser de esos voltajes sea producir una sintaxis.

Pienso que a nadie se le ocurre afirmar que, en el circuito de una radio, existe un homúnculo que es el encargado de interpretar las ondas electromagnéticas como palabras emitidas por el locutor. Del mismo modo, pienso que no se puede hablar de un homúnculo que interpreta los voltajes internos de un ordenador como ceros y unos de un lenguaje computacional, incluso hasta para eso hace falta un homúnculo externo. Lo que hace el diseñador es aprovechar los fenómenos físicos de los componentes electrónicos para implementar con ellos una determinada función que es programable, en el caso de los ordenadores, por contar éstos con dispositivos electrónicos en los que se puede almacenar información. Pero eso no quiere decir que el dispositivo realice esa función, sino que el diseñador aprovecha esa función que el circuito realiza para implementar con ella la función que a él le interesa. Hay por lo tanto una traducción entre lo que el circuito hace y lo que el diseñador quiere, traducción que no realiza el circuito sino el diseñador.

Pero entonces parece que afirmo que el circuito tiene un lenguaje, pienso que no he hecho tal cosa ya que, al hablar de que el circuito realiza una función, no pretendo decir que el circuito ejecuta, como algo propio, un proceso con un fin determinado. Incluso cuando hablo de que el circuito realiza una función necesito de un homúnculo externo que interprete lo que hace el circuito como una función, el circuito no interpreta nada, sólo da unos determinados voltajes, que el homúnculo externo llama “resultado”, en unos lugares del circuito, a los que el homúnculo externo da el nombre de “salida del circuito”, que son consecuencia de los diversos fenómenos físicos que los componentes del circuito producen al circular por ellos las corrientes que se generan al situar en puntos del circuito, que el homúnculo externo llama “entrada del circuito”, unos voltajes, que el homúnculo externo llama “entrada”. Cualquier interpretación que intente situarse por encima de esa explicación física necesariamente requerirá una traducción, y por tanto de un homúnculo externo que la realice.

La objeción que inmediatamente pondrán los defensores del punto de vista de la tercera persona es decir que eso mismo pasa con el hombre y no por eso consideramos que carece de sintaxis, sino que es el sistema como

un todo el que genera la sintaxis y la semántica. Searle responderá hablando de que esa traducción, en el caso del hombre, es intrínseca, ya que en los niveles superiores el punto de vista de la primera persona permite descubrir que en el homúnculo es intrínseco, pero me parece que la respuesta debe contestarse con más detenimiento. Pienso que Searle considera que las neuronas poseen esa capacidad intrínseca, ya que la visión unitaria que Searle tiene del ser humano le impide hablar de dos realidades en el hombre, una que ejecuta un proceso y otra que lo interpreta. Pero con esto no quiere decir que las neuronas traduzcan aquello que están haciendo en una función, sino que, cuando las neuronas actúan, la acción que realizan tiene un fin que coincide con el fin de la acción de ese ser humano visto como un sistema, como una totalidad: fenómeno que Searle no es capaz de descubrir en las acciones de un ordenador.

Otra de las consecuencias que podemos obtener es que la interpretación computacional adolece de poderes causales. Con esto lo que Searle quiere destacar, es el carácter intrínseco de los fenómenos mentales, que hace que aunque yo pueda descubrir cuáles son los patrones de los procesos que se producen en mi cerebro, no basta con su implementación, mediante un sistema físico con los mismo patrones, para que ese sistema tenga los poderes causales que tiene mi cerebro. Por ejemplo, al decir que un ordenador controla el vuelo de un avión, como es en el caso cuando se conecta el piloto automático, yo no puedo dar una explicación de cómo lo hace basándome solamente en los patrones del circuito del piloto automático, necesito asignarle una interpretación computacional (lo que Dennett llamaba estudiarlo con una “actitud” determinada), es decir, no intrínseca al sistema, para que los patrones del circuito del piloto automático tengan poderes causales para controlar el avión. Es cierto que «existe una clase de equivalencia de los sistemas físicos que nos permite codificar la información de las pautas que se dan en los sistemas mediante características físicas que son intrínsecas al sistema, como pueden ser los niveles de voltaje, (...) pero las interpretaciones (que hacemos de estas características físicas) siguen siendo no intrínsecas al sistema»<sup>164</sup>. Pero es posible suponer que se pueden descubrir esas pautas en el cerebro, de forma que lo que necesitara la ciencia cognitiva computacional fuera el acaecimiento de estas pautas intrínsecas. «Pero (...) el descubrimiento de las pautas no explicaría todavía lo que estamos intentando explicar. No estamos intentando descubrir cómo un homúnculo externo puede asignarnos una interpretación computacional a los procesos del cerebro. Más bien, intentamos explicar la ocurrencia de determinados fenómenos biológicos como son la comprensión consciente de una frase, o la experiencia visual consciente de una escena. Esta explicación necesita una comprensión de los procesos físicos, en bruto, que pro-

ducen el fenómeno»<sup>165</sup>. Es decir, lo que Searle busca destacar es que lo que intenta comprender no es cómo puedo asignar a un sistema neuronal una interpretación computacional, sino cómo ese sistema neuronal es capaz de actuar intentando alcanzar el fin que el ser humano, como un todo, busca mediante esa acción.

Otra consecuencia que destaca Searle es la falsedad de la tesis que afirma que el cerebro procesa información. Esta tesis sostiene que los procesos mediante los cuales el cerebro procesa información son los mismos procesos que realiza el ordenador para procesar la misma información. Pero basta con comparar lo que sucede en un ordenador con lo que ocurre en el cerebro para percibir que está equivocada. «En el caso del ordenador es un agente externo el que codifica la información en una forma que pueda ser procesada por los circuitos del ordenador. (...) El ordenador a continuación pasa a través de una serie de estados eléctricos diferentes que el observador externo puede interpretar tanto sintáctica como semánticamente aunque, por supuesto, el *hardware* no posee una semántica o sintaxis intrínseca: todo se encuentra en el ojo del observador. (...) En el caso del cerebro, ninguno de los procesos relevantes que se producen son relativos al observador»<sup>166</sup>. Es decir, cualquier fenómeno cerebral es tan concreto como un huracán o como la digestión de una comida, y sabemos que «podemos hacer, con el ordenador, un modelo de procesamiento de la información de ese evento o de su producción, del mismo modo que podemos hacerlo del tiempo, de la digestión, o de cualquier otro fenómeno, pero los fenómenos no son en sí mismos, por eso, sistemas de procesamiento de la información. (...) Confundir estos sucesos y procesos con la manipulación formal de símbolos, es confundir la realidad con el modelo»<sup>167</sup>. Searle insiste en la misma idea pero bajo un nuevo aspecto, ahora lo que quiere destacar es que ese carácter intrínseco del proceso neuronal impide que lo consideremos como un procesamiento de la información, ya que el fenómeno neuronal es un fenómeno independiente de la interpretación que le queramos dar. Lo que quiere resaltar Searle es que el mismo fenómeno neuronal no puede tener interpretaciones distintas, en cambio el mismo procesado de la información puede implementar funciones distintas. Por lo tanto, interpretar los fenómenos mentales como procesamiento de la información sólo es una metáfora válida si lo que se considera es una interpretación externa del fenómeno, pero esa metáfora también puede ser aplicada a la digestión o a un huracán, ya que ambos fenómenos reciben unos elementos de entrada y producen unos resultados que pueden ser explicados como un procesado de la información. Cualquier otro intento de dar a esta metáfora un alcance mayor, es decir, más cercano al fenómeno real que se produce en el cerebro, es un intento desviado.



Por tanto, «la pregunta “¿Es el cerebro un ordenador digital?” está mal definida. Si preguntase “¿Podemos asignar una interpretación computacional al cerebro?” la respuesta trivial sería que sí, ya que podemos asignar una interpretación computacional a cualquier cosa. Si en cambio preguntase “¿Son los procesos del cerebro intrínsecamente computacionales?” la respuesta trivial sería que no, ya que nada es intrínsecamente computacional, excepto, claro está, los agentes conscientes que realizan intencionalmente computaciones»<sup>168</sup>. Pero, ¿cómo puedo distinguir entre algo que posee intencionalmente computaciones y algo que no lo posee? La respuesta no es otra que la que Searle lleva defendiendo desde el principio, esta diferenciación sólo se puede lograr cuando se adopta un punto de vista de la primera persona.

### 3. RESUMEN

Putnam pretende mostrar que no es posible lograr una teoría reduccionista del significado. Para ello señala en primer lugar dos características del significado, que son su dependencia respecto al entorno y lo que Putnam llama la división del trabajo lingüístico, que hacen imposible aceptar una teoría que no tenga en cuenta esta realidad. La forma más fácil de rechazar esta argumentación es la de negar esas dos características, pero la mayoría de los filósofos han terminado aceptando estos dos argumentos. Es necesario, por tanto, concebir un nuevo tipo de representación que no vaya en contra de ninguno de los argumentos de Putnam. Surge así la idea del contenido estricto, que pretende considerar dos tipos de representaciones, unas que son las que Putnam considera, y otras que serían lo que hay de común entre las representaciones de ese término, pero no como contenido, ya que si no seguiría yendo en contra de los argumentos de Putnam, sino como contenido de una forma potencial, esta es la tesis de Fodor, o como papel funcional, que es la de Block. Pero este intento acaba teniendo que dar algún contenido perceptual a ese contenido estricto, y ahí vuelven a surgir los problemas que Putnam señalaba.

Pero en un principio esto no ha mostrado que no sea posible conseguir reducir el significado con teorías que no sean representacionistas. Para mostrar esta irreductibilidad Putnam utiliza tres argumentos que son el holismo del lenguaje, su normatividad y que la evolución no es capaz de generar todos los términos que se utilizan. Si unimos estos tres argumentos, cualquier intento de proponer reducir el significado contradice alguno de ellos. Por tanto, o rechazamos alguno de los argumentos o tendremos que aceptar que el significado es irreductible.



Una posible solución vuelve a ser alejarnos todavía más del contenido, e intentar explicar el significado mediante los procesos mentales. Putnam considera también esta alternativa, ya que fue él mismo quien la propuso, para mostrar que el proceso mental que estas teorías proponen no existe, y que por lo tanto el funcionalismo no funciona. Para mostrar esta tesis comienza a considerar las posibles soluciones a los problemas que planteaba el sostener una teoría representacionista, y así una primera solución sería incluir los aspectos del entorno en la noción de estado computacional, pero es que el entorno no son sólo las cosas externas, sino que también incluye un criterio de razonabilidad a la hora de interpretar lo que una persona dice. Pero yo puedo proponer como solución el que ese criterio de razonabilidad, en principio, pueda ser formalizado.

Pero ese intento de formalización volvería a encontrarse con un nuevo problema, que es el de extraer, de su propio modelo, una noción de sinonimia. La única solución posible es considerar un stock de palabras universales y biológicamente innatas a partir de las cuales se definen las demás, pero esto no es posible por el holismo del significado.

Hay que seguir buscando una noción de sinonimia que posea el menor contenido posible y que sea más bien una relación computable. Pero esta solución sigue teniendo problemas, ya que es difícil, si no imposible, hacer una descripción completa del entorno empleando sólo un conjunto estandarizado de parámetros. A esta dificultad hay que añadir el que la teoría propuesta, acerca de la sinonimia, debe lograr considerar todas las teorías posibles. Esto un hombre no es capaz de hacerlo, por tanto mucho menos podrá formalizarlo en una teoría. Pero el que no pueda realizarlo no implica que no pueda hacerse. Lo que ocurre es que los esquemas conceptuales que, en principio, deben ser interpretados son infinitos, por lo tanto su única formalización posible es una lista infinita, y aceptar esta solución lleva consigo el reconocer que el fenómeno en estudio es irreducible.

La única solución que queda abierta es la de rechazar el formalismo computacional, y es aquí donde surge la psicología popular que propone que existe un estado cerebral para cada actitud proposicional. Pero el problema de esta teoría es que no consigue determinar esos estados mediante nociones a las que puedan reducirse todas las nociones posibles.

Una vez que Putnam ha mostrado que la intencionalidad es primitiva e irreducible, dirige su mirada a la única teoría que puede no tener en cuenta estas dos propiedades de la intencionalidad, que son aquellas teorías que proponen que la intencionalidad no existe, es decir, que es eliminable. El problema que tiene eliminar la intencionalidad, es que al final acaba eli-

minando también la verdad y la referencia. Algunos propusieron una teoría que intentaba dar razón de la verdad y la referencia sin tener que recurrir a la intencionalidad, es lo que se conoce como teoría desentrecomilladora de la verdad. El problema de esta teoría está, por un lado, en que aunque sí da razón de la verdad de una oración en un lenguaje determinado, en cambio no sabe dar razón de la diferencia de significado de la misma oración en dos lenguajes distintos, y por otro, en que o se elimina el realismo o se acepta la intencionalidad, pero en este segundo caso uno tendría que dejar de ser eliminativista también respecto a las actitudes proposicionales. La única postura coherente para el eliminativista es ser irrealista, pero esa posición no resulta la más convincente.

Searle, por su parte, realiza una crítica de las teorías reduccionistas de la mente por considerar que dejan de lado un aspecto importante de los fenómenos mentales que es la subjetividad. Desde el momento en que una teoría acepta como premisa que la única forma objetiva de estudiar un problema es haciendo uso de un punto de vista de la tercera persona, esa teoría está necesariamente abocada a considerar que algo es inteligente si se comporta como si lo fuera, ya que mi único medio de observación es lo objetivo.

La primera consecuencia de adoptar este punto de vista de la tercera persona es que se hace difícil explicar que realidades que siempre han sido aceptadas como no inteligentes lo sigan siendo al aplicar este criterio. O también se puede uno ir al extremo opuesto y terminar eliminando la conciencia, ya que como no es posible observarla objetivamente, entonces, tal vez, convenga negarla. Yo puedo dedicarme a observar a la otra persona buscando descubrir su conciencia, pero qué me permite afirmar que dentro de ella se están produciendo procesos que son conscientes. Desde una visión objetiva sólo me es válido su comportamiento, y entonces o considero que todo lo que se comporta inteligentemente tiene conciencia, o explico que el ser humano no tiene conciencia o que la conciencia no es más que un fenómeno neuronal que me informa de los fenómenos neuronales que se producen en mi cerebro.

Searle no quiere afirmar que esté en contra de la reducción como método científico, sino que el reduccionismo es un método epistemológicamente objetivo que sólo considera aquello que puede ser estudiado objetivamente. En el caso de los fenómenos mentales lo que los caracteriza es que son fenómenos subjetivos, ya que lo importante no es cómo me comporto, sino por qué me comporto de esa manera. Pero esa respuesta exige un punto de vista que considere mi subjetividad ontológica, es decir, que mis fenómenos mentales son míos, por tanto, cualquiera que pretenda eliminar esa

subjetividad estará dejando de lado justo aquello a lo que se debería dedicar todo su interés.

Para mostrar la necesidad de emplear este punto de vista de la primera persona Searle ideó lo que se conoce como el experimento de la habitación china, con la que pretende hacer ver que aunque el comportamiento final sea el mismo, es de sentido común que mientras un chino entiende el chino, en cambio el hombre que está dentro de la habitación no entiende el chino, por tanto lo realmente decisivo no es el comportamiento, sino que hay algo más que es necesario considerar. Los reduccionistas y eliminativistas afirmarán que es la habitación como un todo la que entiende chino, y son así coherentes con el punto de vista de la tercera persona. Pero entender una lengua es algo más que manipular unas reglas, es querer significar algo, pretender hacer referencia a una realidad que está más allá del propio lenguaje y de las reglas con que se construyen las oraciones de ese lenguaje, y aunque se tome a la habitación como un todo o se internalice todo el sistema en el hombre que está dentro de la habitación, no por eso el sistema sabe a qué está haciendo referencia. Es como afirmar que el sistema de calefacción, considerado como un todo, “sabe” cuando la casa está suficientemente caliente y por tanto deja de calentarla.

Pero entonces Searle parece que defiende un *asimetría epistemológica* a la hora de estudiar los fenómenos mentales. Así es si no entendemos por asimetría epistemológica la posesión de un sentido interno que me permita contemplar mis estados mentales o el acceso privilegiado a mis estados intencionales como si fuera una habitación en la que yo sólo puedo entrar. Cuando Searle acepta una asimetría, ésta es debida a que la realización de actos del habla -y el significar cosas con expresiones- se produce a un nivel de intencionalidad intrínseca de primera persona, y por lo tanto sólo desde un punto de vista de la primera persona puede afirmarse que existe tal intencionalidad. Esto hace que el emplear un punto de vista de la primera persona sea necesario para no dejar de lado los contenidos intencionales, los que han intentado objetivizar el estudio del significado han cometido el error de olvidar estos contenidos y eso les ha llevado a una mala comprensión del lenguaje.

Si se aplican estas consideraciones al intento de lograr comportamientos inteligentes siguiendo las propuestas de las teorías computacionales de la mente, se descubren algunas características que al menos deberían de hacer pensar que hay algo en la teoría que no cuadra. La primera característica que sorprende es la múltiple posibilidad de implementación de todo proceso computacional, que tiene dos graves consecuencias. La primera es que esa múltiple realizabilidad de una función se logra gracias a que esa



interpretación como una función, de los fenómenos físicos que se producen en el ordenador, está realizada por un observador externo. Por tanto, cualquier cosa que pudiera ser interpretada como realización de un proceso computacional sería un ordenador, entonces todo puede ser un ordenador ya que a todos los fenómenos físicos puedo darles una interpretación de tipo computacional. Pero esta universalidad de la aplicación de interpretaciones computacionales se puede resolver fácilmente mediante la restricción de la extensión del término computación, dotándolo de unas características, (programación, poderes causales, etc.), que no lo hagan aplicable a todas las realidades físicas.

La segunda consecuencia es que se descubre que la sintaxis no es intrínseca a la física, lo cual comporta un conjunto de desastrosas consecuencias. Si el experimento de la habitación china había mostrado no sólo la necesidad de adoptar un punto de vista de la primera persona al estudiar el problema de la mente, sino también que la semántica es extrínseca a la sintaxis, ahora con esta afirmación Searle va todavía más lejos y afirma que, a su vez, la sintaxis, en los ordenadores y en todo aquello que comporta una interpretación computacional, es extrínseca a lo físico. Que la sintaxis sea extrínseca a la física tiene como primera consecuencia la insuficiencia de la eliminación del homúnculo mediante un descenso recursivo. Si yo acepto que en los niveles superiores existe un observador externo que es el que interpreta lo que hace el ordenador como una determinada función, entonces he de aceptar que necesito también un observador externo para dar razón de la sintaxis, ya que la sintaxis no es intrínseca a lo físico. Una segunda consecuencia es que la interpretación computacional adolece de poderes causales, y que por tanto no basta con que se logre implementar un sistema físico con los mismos patrones que los de mi sistema neuronal para que ese sistema tenga los poderes causales que tiene mi cerebro, ya que esos poderes causales los tiene el sistema como propios y no dependen de la interpretación que yo haga de ellos. Una última consecuencia es la falsedad de la tesis que afirma que el cerebro procesa información ya que ese carácter intrínseco del proceso neuronal impide que se le pueda considerar como un procesamiento de la información, debido a que el fenómeno neuronal es un fenómeno independiente de la interpretación que le queramos dar. Esto se puede descubrir fácilmente al considerar que el mismo fenómeno neuronal no puede tener interpretaciones distintas y en cambio el mismo procesado de la información puede implementar funciones distintas.







## CITAS BIBLIOGRAFICAS

1. PUTNAM, H., *Representation and Reality*.
2. SEARLE, J. R., *The Rediscovery of the Mind*.
3. PUTNAM, H., *Representation and Reality*, p. 108. En esta cita no sigo la traducción de G. Ventureira.
4. Afirmar que la intencionalidad es un fenómeno primitivo es enunciar positivamente la llamada "tesis de Brentano" que afirma que "la intencionalidad no será reducida ni tampoco eliminada".
5. Cfr. TALBOT, M., "Review of H. Putnam's *Representation and Reality*", p. 453; SEARLE, J. R., *The Rediscovery of the Mind*, p. XIII.
6. En palabras del propio Putnam: «Primitive just means "not reducible to nonintentional notions"» (PUTNAM, H., *Representation and Reality*, p.110).
7. PUTNAM, H., *Representation and Reality*, p. 19.
8. ARISTOTELES, *De interpretatione*, Valencia, Teorema, trad. esp. de A. García Suárez y J. Velarde Lombraña, 1977, 16a 3-8.
9. «So we can say that Aristotle seems to have regarded the human soul quasi as a photographic plate, on which the things leave their pictures» (ARENS, H., *Aristotle's Theory of Language and its Tradition*, Amsterdam, John Benjamins, 1984, p. 30).
10. Cfr. BYNUM, T. W., "A New Look at Aristotle's Theory of Perception", en M. Durrant (ed.) *Aristotle's De Anima in Focus*, London, Routledge, 1993, p. 102.
11. ARISTOTELES, *De Interpretatione*, 16a 7-8.
12. Cfr. ARENS, H., *Aristotle's Theory of Language and its Tradition*, pp. 28-29.
13. Cfr. SCHNEIDER, J., "Omoioima", en Gerhard Kittel y Gerhard Friedrich (eds.) *Grande Lessico del Nuovo Testamento*, vol. VIII, Brescia, Paideia, trad. ital. de F. Ronchi, G. Casanova y benedictinos del monasterio de Santa Escolástica, 1972, cols. 535-536.
14. Cfr. PLATON, "Phaedrus", en *Platonis Opera*, Vol. II, Oxford, Typographeo Clarendoniano, 1964, 250b.

15. Cfr. BARREAU, H., *Aristóteles y el análisis del saber*, Madrid, EDAF, trad. esp. de I. Gómez Romero, 1977, p. 33.
16. Cfr. ARISTOTELES, *Acerca del Alma*, 431a 15-17 y 431b 2.
17. Cfr. PUTNAM, H., *Representation and Reality*, pp. 20-21.
18. *Ibidem*, pp. 21-22.
19. PUTNAM, H., *Representation and Reality*, p. 20.
20. Cfr. *Ibidem*, pp. 23-36. Esta argumentación la propuso por primera vez en "The Meaning of 'Meaning'", en *Mind, Language, and Reality: Philosophical Papers*, vol. II, Cambridge, Cambridge University Press, 1975.
21. PUTNAM, H., *Representation and Reality*, p. 24.
22. «Language is a form of cooperative activity, not an essentially individualistic activity. Part of what is wrong with the Aristotelian picture is that it suggests that everything that is necessary for use of language is stored in each individual mind; but no actual language works that way» (*Ibidem*, p. 25). Como se observa, Putnam no habla de la teoría representacionista, sino de la teoría aristotélica, que he preferido traducir por teoría representacionista para evitar los problemas de atribuir esta tesis a Aristóteles.
23. *Ibidem*, p. 25.
24. PUTNAM, H., *Representation and Reality*, p. 36.
25. Fodor ha propuesto posteriormente una teoría un poco distinta en su artículo "Banish DisContent", en J. Butterfield (ed.) *Language, Mind and Logic*, Cambridge, MA, Cambridge University Press, 1986. Pero aunque habla de una función del contexto al referente, esta función sigue estando en la cabeza, y posee los mismos inconvenientes que Putnam señala a continuación para la teoría de Searle.
26. Putnam afirma que así lo admitió en un debate que siguió a la conferencia "Why Meanings aren't in the Head" que Putnam dió en la Universidad de Rutgers el 13 de marzo de 1986 (Cfr. PUTNAM, H., *Representation and Reality*, p. 26, n. 7).
27. PUTNAM, H., *Representation and Reality*, p. 28.
28. *Ibidem*, pp. 28-29.
29. PUTNAM, H., *Representation and Reality*, p. 30.
30. *Ibidem*, p. 43.
31. PUTNAM, H., *Representation and Reality*, p. 45.
32. *Ibidem*, p. 46.
33. *Ibidem*, p. 46.
34. Cfr. FODOR, J., *Psicosemántica*, Madrid, Tecnos, trad. esp. de O. L. González-Castán, 1994; "Banish DisContent".

35. Putnam conoce estas dos publicaciones de Fodor, y las cita en la nota 19 del capítulo 2 y en la 1 del capítulo 3. Si no les dedica una mayor atención es porque considera que esta nueva propuesta sigue sin resolver los problemas que ya tenía su propuesta anterior.
36. FODOR, J., *Psicosemántica*, p. 84.
37. *Ibidem*, p. 81.
38. FODOR, J., *Psicosemántica*, p. 82.
39. Cfr. FODOR, J., "A Modal Argument for Narrow Content", en *The Journal of Philosophy*, LXXXVIII, 1991, pp. 7-9.
40. *Ibidem*, p. 8.
41. Forma de llamar a XYZ en la Tierra Gemela.
42. FODOR, J., "A Modal Argument for Narrow Content", p. 25.
43. FODOR, J., "A Modal Argument for Narrow Content", p. 21.
44. FODOR, J., *Psicosemántica*, p. 63.
45. *Ibidem*, p. 62.
46. Cfr. BERNIER, P., "Narrow Content, Context of Thought and Asymmetric Dependency", en *Mind and Language*, VIII, 1993, pp. 327-340.
47. Cfr. BLOCK, N., "Functional Role and Truth Conditions", en *Proceedings of the Aristotelian Society*, Sup. LXI, 1987, pp. 157-181.
48. PUTNAM, H., *Representation and Reality*, p. 47.
49. *Ibidem*, p. 52.
50. BEATTIE, J., "Is Putnam Inconsistent and Parochial?", en *Australasian Journal of Philosophy*, LXXI, 1993, pp. 316-321.
51. MCGINN, C., *Mental Content*, Oxford, Blackwell, 1989, pp. 7-8.
52. Cfr. *Ibidem*, pp. 14-22, 42-43.
53. OWENS, D., "Review of C. McGinn's *Mental Content*", en *Mind*, IC, 1990, p. 115.
54. Cfr. PUTNAM, H., *Representation and Reality*, p. XIII.
55. Cfr. *Ibidem*, pp. 8-15.
56. QUINE, W. V., "Two Dogmas of Empiricism", en *Philosophical Review*, LX, 1951, pp. 20-43; *Palabra y objeto*, Barcelona, Labor, trad. esp. de M. Sacristán, 1968.
57. PUTNAM, H., *Representation and Reality*, p. 9.
58. *Ibidem*, p. 13.
59. *Ibidem*, p. 15.
60. Cfr. PUTNAM, H., *Representation and Reality*, pp. 15-18.



61. Cfr. PUTNAM, H., *Representation and Reality*, p. 74.
62. PUTNAM, H., *Representation and Reality*, p. 75.
63. *Ibidem*, p. 75.
64. *Ibidem*, p. 75.
65. Cfr. PUTNAM, H., *Representation and Reality*, p. 79.
66. Cfr. *Ibidem*, p. 81.
67. PUTNAM, H., *Representation and Reality*, p. 81.
68. Cfr. QUINE, W. V. O., *Palabra y objeto*, pp. 41-53.
69. QUINE, W. V. O., *Palabra y objeto*, p. 51.
70. PUTNAM, H., *Representation and Reality*, p. 53.
71. *Ibidem*, p. 85.
72. *Ibidem*, p. 87.
73. PUTNAM, H., *Representation and Reality*, p. 87.
74. *Ibidem*, p. 89.
75. PUTNAM, H., *Representation and Reality*, p. 92.
76. PUTNAM, H., *Representation and Reality*, p. 93.
77. *Ibidem*, p. 94.
78. *Ibidem*, p. 95.
79. PUTNAM, H., *Representation and Reality*, p. 96.
80. *Ibidem*, p. 125.
81. «On the notion of "similarity of possible worlds" and on a theory of "events" which in turn presupposes his "theory of universals"» (PUTNAM, H., *Representation and Reality*, p. 96). «Lewis ha comparado la manera en que esta teoría se presenta con una historieta que un detective podría contar sobre la muerte de un tal Mr. Body: "X, Y y Z han conspirado para asesinar a Mr. Body. Hace diecisiete años, en las minas de oro de Uganda, X era socio de Body... La semana pasada Y y Z se reunieron en un bar de Reading... El jueves por la noche, a las 11.17, Y fue al ático y colocó una bomba de relojería... Diecisiete minutos más tarde, X se encontró con Z en la sala de billar y le dio una pistola cargada... Justamente cuando la bomba estalló en el ático, X disparó tres tiros hacia el estudio a través de los ventanales franceses" (LEWIS, D. K., "Psychophysical and Theoretical Identifications", p. 208). Lo mismo que podemos seguir esta historieta sin saber quiénes son X, Y y Z, Lewis mantiene que podemos seguir el discurso ordinario sobre las relaciones causales de los eventos mentales sin saber qué procesos neurales originan» (BECHTEL, W., *Filosofía de la mente*, p. 153).
82. PUTNAM, H., *Representation and Reality*, p. 98.

83. *Ibidem* p. 101.
84. Las "trivialidades" de la psicología popular son, para Lewis, aquellas cosas que todos sabemos acerca de los estados mentales, es decir, que forman parte de nuestro conocimiento común, y que constituyen una definición implícita de los estados mentales.
85. PUTNAM, H., *Representation and Reality*, p. 102.
86. PUTNAM, H., *Representation and Reality*, pp. 103-105.
87. Cfr. TALBOT, M., "Review of H. Putnam's *Representation and Reality*", p. 454.
88. Cfr. PUTNAM, H., *Representation and Reality*, pp. 59-60.
89. *Ibidem*, p. 60.
90. Cfr. TARSKI, A., "The Concept of Truth in Formalized Languages", en *Logic, Semantics, Metamathematics*, Oxford, Clarendon Press, 1956.
91. PUTNAM, H., *Representation and Reality*, p. 61.
92. TARSKI, A., "The Concept of Truth in Formalized Languages", pp. 165-166.
93. Cfr. TARSKI, A., "The Concept of Truth in Formalized Languages", pp. 170-173.
94. FIELD, H., "Tarski's Theory of Truth", en *The Journal of Philosophy*, LXIX, 1972, pp. 347-375.
95. DAVIDSON, D., *De la verdad y de la interpretación*, Barcelona, Gedisa, trad. esp. de G. Filippi, 1990, p. 222.
96. PUTNAM, H., *Representation and Reality*, p. 66.
97. DAVIDSON, D., *De la verdad y de la interpretación*, p. 208.
98. DAVIDSON, D., *De la verdad y de la interpretación*, p. 209.
99. Esto queda mostrado de forma patente en el tipo de definición que, para un lenguaje que consta de dos oraciones, nos puede dar esta teoría:  
 «L1 = df the language *L* such that, for any term *W* and for any *x*, {(*W* = "Der Mond" and *x* = the moon) or (*W* = "Schnee" and *x* = the substance snow) or (*W* = "blau" and *x* is blue) or (*W* = "weiss" and *x* is white)} if and only if (case i) *W* = "Der Mond" and *x* = the moon, or (case ii) *W* = "Schnee" and *x* = the substance snow, or (case iii) *W* = "blau" and *x* is blue, or (case iv) *W* = "weiss" and *x* is white; and such that for any sentence *S*, {(*S* is spelled "Der Mond ist blau" and the moon is blue) or (*S* is spelled "Schnee ist weiss" and the snow is white)} if and only if (case a) *S* is spelled "Der Mond ist blau" and the moon is blue, or (case b) *S* is spelled "Schnee ist weiss" and the snow is white; and (syntactic restriction) no inscription with any other spelling is a well formed formula of *L* » (PUTNAM, H., *Representation and Reality*, p. 65), ya que un lenguaje L2 en el que las dos oraciones tengan otro significado satisface esta definición.

100. RORTY, R., *La filosofía y el espejo de la naturaleza*, Madrid, Cátedra, trad. esp. de J. Fernández Zulaica, 1983, p. 274.
101. PUTNAM, H., *Representation and Reality*, p. 68.
102. PUTNAM, H., *Representation and Reality*, p. 69.
103. HOUGHTON, D., "Rorty's Talk-About", en A. Malachowski (ed.) *Reading Rorty*, Oxford, Blackwell, 1990, p. 161.
104. Es el caso de Rorty. «Rorty was at one time an eliminationist in the philosophy of mind, and has gone on to become an eliminationist with respect to truth and reference (or with respect to the idea that there are "interesting" notions of truth and reference)—as a result of which he has also gone on to eliminate the whole problematic of scientific realism!» (PUTNAM, H., *Representation and Reality*, p. 70).
105. *Ibidem*, p. 70.
106. SEARLE, J. R., *The Rediscovery of the Mind*, p. 16.
107. SEARLE, J. R., *The Rediscovery of the Mind*, p. 94.
108. SEARLE, J. R., *Intencionalidad*, Madrid, Tecnos, trad. esp. de E. Ujaldón Benítez, 1983, p. 84.
109. SEARLE, J. R., *The Rediscovery of the Mind*, p. 97.
110. BURGE, T., "Philosophy of Language and Mind: 1950-1990", pp. 31-32.
111. SEARLE, J. R., *The Rediscovery of the Mind*, p. 17.
112. SEARLE, J. R., *The Rediscovery of the Mind*, p. 97.
113. Cfr. SEARLE, J. R., *Mentes, cerebros y ciencia*, Madrid, Cátedra, trad. esp. de L. Valdés, 1990, pp. 19-21.
114. SEARLE, J. R., *The Rediscovery of the Mind*, pp. 55-56.
115. SEARLE, J. R., *The Rediscovery of the Mind*, p. 132.
116. *Ibidem*, p. 132.
117. SEARLE, J. R., *The Rediscovery of the Mind*, pp. 99-100.
118. NAGEL, T., "What Is It Like to Be a Bat?", en Ned Block (ed.) *Readings in Philosophy of Psychology*, vol. I, London, Methuen, 1980, p. 160.
119. Cfr. NAGEL, T., "What Is It Like to Be a Bat?"; KRIPKE, S. A., "Naming and Necessity" en D. Davidson y G. Harman (eds.) *Semantics of Natural Language*, Dordrecht, Reidel, 1971, pp. 253-355 y 763-769; JACKSON, F., "Ephiphenomenal Qualia," en *Philosophical Quarterly*, XXXII, 1982, pp. 127-136.
120. SEARLE, J. R., *The Rediscovery of the Mind*, pp. 117-118.
121. SEARLE, J. R., *The Rediscovery of the Mind*, p. 120.
122. *Ibidem*, p. 121.



123. SEARLE, J. R., *The Rediscovery of the Mind*, pp. 121-122.
124. *Ibidem*, p. 123.
125. SEARLE, J. R., *The Rediscovery of the Mind*, p. 15.
126. Cfr. *Ibidem*, p. 24.
127. «According to the atomic theory of matter, the universe consists entirely of extremely small physical phenomena that we find it convenient, though not entirely accurate, to call 'particles'» (*Ibidem*, p. 86).
128. SEARLE, J. R., *The Rediscovery of the Mind*, p. 105.
129. SEARLE, J. R., *The Rediscovery of the Mind*, p. 126.
130. *Ibidem*, p. 121.
131. Cfr. SEARLE, J. R., "Minds, Brains, and Programs", en *Behavioral and Brain Sciences*, III, 1980, pp. 417-424.
132. Cfr. DRETSKE, F., "Machines and the Mental", en *Proceedings and Addresses of the American Philosophical Association*, LIX, 1985-86, pp. 23-33. Propone en este artículo el estudio de la operación aritmética de la suma, y descubre que la diferencia entre la suma realizada por una persona y la que realiza la máquina es que mientras para la primera la operación que realiza está dotada de significado, en cambio para la máquina ese significado le es asignado y por lo tanto no le es propio.
133. DENNETT, D. C., "The Milk of Human Intentionality", en J. L. Garfield (ed.) *Foundations of Cognitive Science*, New York, Paragon House, 1990, pp. 212-216.
134. *Ibidem*, p. 213.
135. *Ibidem*, p. 213.
136. SEARLE, J. R., *Mentes, cerebros y ciencia*, p. 40.
137. Cfr. HAYES, P., HARNARD, S., PERLIS, D. y BLOCK, N., "Virtual Symposium on Virtual Mind", en *Mind and Machines*, II, 1992, pp. 217-238. Este artículo es bastante interesante ya que propone muchas más objeciones al experimento de la habitación china. Pienso que el error que cometen, al considerar el argumento de Searle, es quedarse en problemas laterales y no entrar de lleno en el auténtico problema que es el de determinar si un ordenador puede tener una subjetividad intrínseca. Por ahora, no conozco ningún ordenador que la tenga y por tanto si se quiere rechazar el argumento de la habitación china, no necesitan otra cosa que hacer un ordenador con esa subjetividad. Pero si se supone que se lograra (cosa que Searle considera posible aunque mi experiencia en este campo me lleva a verlo como irrealizable), entonces surge un nuevo problema: ¿Cómo sé que esa subjetividad es intrínseca? La podría presuponer, pero nunca podría afirmar que la posee.
138. HAYES, P., HARNARD, S., PERLIS, D. y BLOCK, N., "Virtual Symposium on Virtual Mind", p. 223.

139. Es cierto que no hay ningún hombrequito dentro del ordenador, pero sí que existe lo que se llama la "unidad de control", en la que se encuentra el procesador de instrucciones del microprocesador. Es un circuito electrónico que lee las instrucciones que están almacenadas en memoria y las traduce en órdenes que envía al resto de la circuitería del microprocesador. Por tanto, sólo puedo aceptar que la metáfora del hombrequito es desviada si lo que se quiere decir con ello es que el ordenador no realiza una operación que pueda ser llamada *lectura*, u *obediencia*, o *seguir* unas leyes, en el sentido de que para realizar tales operaciones es necesario entender lo que se está haciendo. Pero si éste fuera el sentido que le da Hayes, entonces estaría yendo a favor de la tesis de Searle. Hayes me parece que lo que quiere defender es que no existe una parte concreta del ordenador que haga de hombrequito, sino que dentro del ordenador lo que se tiene son corrientes que van produciendo, mediante esa circuitería electrónica, la ejecución del proceso que corresponde a la función cognitiva que ese ordenador implementa. Pienso que no es acertado afirmar que basta esa ilocalizabilidad del hombrequito en la circuitería para rechazar el experimento de la habitación china. Para poder desecharlo es necesario hacer ver dónde logra el ordenador entender lo que hace. Hayes probablemente considera que esto puede descubrirse al estudiar su comportamiento inteligente, considerándolo como un todo, pero entonces el termostato también sería inteligente. Pienso, en cambio, que la metáfora del hombrequito es bastante correcta, y por eso al explicar la unidad de control suele usarse esa imagen, ya que lo que se quiere indicar con esa metáfora es que el procesador de instrucciones no "sabe", (no entiende), lo que está haciendo. Unas corrientes que circulan por unos circuitos electrónicos no pueden producir tal capacidad, ya que si no mi televisión o mi compact-disc también tendrían que tener inteligencia.
140. SEARLE, J. R., "Indeterminacy, Empiricism, and the First Person", en *The Journal of Philosophy*, LXXXIV, 1987, p. 143.
141. *Ibidem*, p. 145.
142. SEARLE, J. R., "Response: Meaning, Intentionality, and Speech Acts", en E. Lepore y R. Van Gulick (eds.) *John Searle and His Critics*, Oxford, Blackwell, 1991, p. 85.
143. SEARLE, J. R., "Meaning, Communication, and Representation", en R. Grandy y R. Warner (eds.) *Philosophical Grounds of Rationality, Intentionality, Categories and Ends*, Oxford, Clarendon Press, 1986, p. 217.
144. Cfr. *Ibidem*, p. 216.
145. SEARLE, J. R., "Indeterminacy, Empiricism, and the First Person", p. 146.
146. TURING, A., "Computing Machinery and Intelligence", en *Mind*, LIX, 1950, pp. 433-460.
147. SEARLE, J. R., *The Rediscovery of the Mind*, p. 202.
148. SEARLE, J. R., *The Rediscovery of the Mind*, p. 205.

149. GARFIELD, J. L., "Introduction", en J. L. Garfield (ed.) *Foundations of Cognitive Science*, New York, Paragon House, 1990, p. xi-xii.
150. SEARLE, J. R., *The Rediscovery of the Mind*, p. 207.
151. *Ibidem*, p. 207.
152. *Ibidem*, p. 209.
153. SEARLE, J. R., *The Rediscovery of the Mind*, pp. 207-208.
154. SEARLE, J. R., *The Rediscovery of the Mind*, p. 209.
155. *Ibidem*, p. 210.
156. Cfr. SEARLE, J. R., "Minds, Brains, and Programs", pp. 417-424.
157. SEARLE, J. R., *The Rediscovery of the Mind*, p. 211.
158. SEARLE, J. R., *The Rediscovery of the Mind*, p. 212.
159. La falacia del homúnculo es considerar que el cerebro es como si fuera un agente que hay dentro de uno y que se usa para ejecutar los programas que producirán los diferentes procesos mentales. Es la metáfora del hombrecito, que antes se vio para los ordenadores, pero utilizada ahora para los seres humanos. El problema de esta metáfora, tanto en el caso del hombre como en el de los ordenadores, es que si se acepta hay que localizar ese homúnculo en el sustrato físico. Para Searle esto no entraña ningún problema en el caso del hombre, ya que éste posee como intrínseca esa capacidad, mientras que en el caso de los ordenadores su homúnculo es la persona que lo está usando, pero ¿quién es el usuario del cerebro? (Cfr. SEARLE, J. R., *The Rediscovery of the Mind*, p. 214).
160. *Ibidem*, pp. 213-214.
161. Cfr. DENNETT, D. C., *Brainstorms: Philosophical Essays on Mind and Psychology*.
162. Un *flip-flop* es un circuito electrónico que se caracteriza por poseer dos estados estables, por eso recibe también el nombre de biestable. Normalmente se hace corresponder sendos estados, en el caso de la electrónica digital, con un 0 ó un 1.
163. SEARLE, J. R., *The Rediscovery of the Mind*, p. 214. Estoy de acuerdo con esta consideración de Searle, ya que cuál es la diferencia entre "multiplicar 6 por 8", "sumar 8 seis" ó "11101110 00000110 00001000". Se puede responder que corresponden a diversos niveles de computación de la misma operación, pero es que esa ascensión sólo se realiza mediante la agrupación de diversas operaciones del nivel inferior en una operación del nivel superior. Por lo tanto, qué poder puede tener esa simple agrupación para convertir en extrínseco algo que, según estos cognitivistas, es intrínseco a lo físico. No significa que neguemos que existan fenómenos en el proceso computacional que sean intrínsecos a lo físico (la intensidad de la corriente que circula por una determinada línea del *bus* de datos, el voltaje en una determinada patilla de un *chip*, etc.), pero es que esos fenómenos



sólo poseerán sintaxis cuando un observador externo se la dé, incluso en los niveles inferiores.

- 164. SEARLE, J. R., *The Rediscovery of the Mind*, p. 222.
- 165. SEARLE, J. R., *The Rediscovery of the Mind*, p. 222.
- 166. SEARLE, J. R., *The Rediscovery of the Mind*, pp. 223-224.
- 167. *Ibidem*, pp. 224-225.
- 168. SEARLE, J. R., *The Rediscovery of the Mind*, p. 225.



## INDICE

PROLOGO .....	65
INDICE DE LA TESIS.....	69
BIBLIOGRAFIA DE LA TESIS.....	71
TABLA DE ABREVIATURAS .....	81
LOS REDUCCIONISMOS CONTEMPORANEOS	
ACERCA DE LA MENTE.....	83
1. LAS CRITICAS DE HILARY PUTNAM.....	83
1.1. La intencionalidad es primitiva.....	84
1.2. La intencionalidad no es reductible .....	94
1.2.1. El significado no es reductible .....	94
1.2.2. El funcionalismo no funciona.....	97
1.3. La intencionalidad es ineliminable .....	102
2. CRITICAS DE JOHN R. SEARLE .....	106
2.1. Los reduccionismos de la mente se basan en una ontología de la tercera persona.....	107
2.2. La mente no es un ordenador digital .....	118
3. RESUMEN .....	125
CITAS BIBLIOGRAFICAS.....	131